TE SHU JU ZHEN

特殊矩阵

SPECIAL MATRICES

陈景良 陳向晖

清华大学出版社

(京)新登字 158 号

图书在版编目(CIP)数据

特殊矩阵/陈景良，陈向晖 编著.—北京：清华大学出版社，2000
ISBN 7-302-04129-6

I. 特… II. ① 陈… ② 陈… III. 矩阵-理论 IV. 0151.21

中国版本图书馆 CIP 数据核字(2000)第 77889 号

出版者：清华大学出版社（北京清华大学学研大厦，邮编 100084）
http://www.tup.tsinghua.edu.cn
印刷者：北京市清华园胶印厂
发行者：新华书店总店北京发行所
开　本：850×1168　1/32　**印张**：25.375　**字数**：635 千字
版　次：2001 年 1 月第 1 版　2001 年 1 月第 1 次印刷
书　号：ISBN 7-302-04129-6 /O・251
印　数：0001～4000
定　价：45.00 元

内 容 提 要

本书是一部全面介绍有专门术语或人名命名的矩阵的论著，无论在学术上还是在应用上都有其独特的作用.

全书共 10 章，内容包括：基础知识，从现代数学的观点阐述了线性代数的基本理论；不可约、对角优势、酉、正规等基本性质矩阵；自伴(Hermite)、正定和半正定等矩阵以及稳定矩阵等；正、非负、循环、素和随机等矩阵，以及M-矩阵和H-矩阵等；Jordan标准形和相似变换、友矩阵和Frobenius矩阵、Schur标准形和奇异值分解、Householder变换和Hessenberg矩阵、Givens变换和 QR分解、Gauss 变换和 LU 分解；带状、轮换、Toeplitz、Hankel、中心对称、同伴和结式等特型矩阵；Kronecker积和Hadamard积等特殊积矩阵，以及各种广义逆矩阵；Jacobi、Gauss-Seidel、SOR、SSOR、AOR 和 SAOR 诸方法的矩阵分裂和迭代矩阵；多项式、非多项式和Hadamard等矩阵函数，以及一般函数矩阵和作为特殊情形的λ-矩阵与有理矩阵；矩阵的有向图，性质 A、相合、幂、整数、奇偶校验、对合、区间和自反等矩阵综述，以及关于素、复对称和自伴等矩阵的进一步的性质.

本书取材丰富，涵盖 280 余种命名矩阵，能反映最新进展；理论严谨，重点突出，择优推证方法；贯穿应用背景或具体应用；结构合理，既有系统性，适合全面阅读，又具可分性，便于选读；灵活实用，查阅方便；深入浅出，阅读本书只需具备微积分和线性代数的基本知识.

本书兼理论专著、工具书、大学有关专业教材或参考书于一体. 读者对象主要为数学及其应用数学和数值数学、工程技术以及经济科学等工作者、大学教师、本科生和研究生.

序　　言

矩阵简史

矩阵并非如同一种容易产生的猜想那样直接源自线性方程组系数的研究.系数阵列导致数学家们发展了**行列式**而不是矩阵.微积分创建合作者 Leibniz 在 1693 年使用了行列式,先于矩阵成为独立研究对象约 150 年.Cramer 在 1750 年建立解线性方程组的行列式基本公式,Gauss 在 1820 年左右提出消去法.这些事件都出现在矩阵概念存在之前.

顺便插一句,Gauss 消去法多年来是作为大地测量学(而不是数学)发展的一部分;称为主元消去法的 Gauss-Jordan 方法最先也是出现在大地测量学手册之中.

矩阵代数得以产生必须具备下述两个条件:

(1) 适当的记号,诸如 a_{ij} 和 A 等;

(2) 矩阵乘法的定义.

很巧合的是这两个紧要因素几乎形成于同一时间,大约 1850 年,而且出自同一国家,英国.除了 Newton 创建微积分外,17、18 至 19 世纪近代数学早期主要成就都是欧洲大陆的数学家们取得的,他们是 Bernoulli,Cauchy, Euler, Gauss 和 Laplace 等.但是到了 19 世纪中期,英国的数学家率先开展各种代数系统基础结构的研究.例如,A. DeMorgan 和 G. Boole 创立了集代数(Boole 代数).

确立矩阵概念和产生"矩阵"一词的动机是试图为研究行列式提供适当的代数语言.1848 年 J. J. Sylvester 引进术语"矩阵",拉丁文为"womb",作为数的阵列的名称.他用"womb"是因为他视矩阵为行

列式的生成体.亦即,矩阵的每 k 行和 k 列的子集(相应行和列确定的子矩阵)生成一个行列式.

在围绕行列式研究而寻求好的记号期间,Sylvester 在 1851 年提议把方形矩阵写成如下形式:

$$\begin{matrix} a_1\alpha_1 & a_1\alpha_2 & \cdots & a_1\alpha_n \\ a_2\alpha_1 & a_2\alpha_2 & \cdots & a_2\alpha_n \\ \vdots & \vdots & \vdots & \vdots \\ a_n\alpha_1 & a_n\alpha_2 & \cdots & a_n\alpha_n \end{matrix} \tag{1}$$

其每个表值(元素)表示成符号之积.他还引入了方阵的缩减记号

$$\begin{pmatrix} a_1 & a_2 & \cdots & a_n \\ \alpha_1 & \alpha_2 & \cdots & \alpha_n \end{pmatrix}, \tag{2}$$

并称诸 a_i 和诸 α_j 为**哑元**(umbra)或**理想元素**.而后, Sylvester 使用这些哑元记号,将(2)的行列式——包括求诸 a_i 匹配诸 α_j 所有置换带符号之积的总和——写成

$$\begin{Bmatrix} a_1 & a_2 & \cdots & a_n \\ \alpha_1 & \alpha_2 & \cdots & \alpha_n \end{Bmatrix}. \tag{3}$$

采用(1)后不久,两种符号 a_i 和 α_j 便被归并成带双下标的一种符号——a_{ij} (Cauchy 在 1812 年实际上已经使用 a_{ij},但其后没有立即被采纳).

矩阵代数起源于 1855 年 A. Cayley 关于**线性变换**的工作.设有线性变换

$$T_1:\ \begin{cases} x' = ax + by \\ y' = cx + dy \end{cases} \quad 和 \quad T_2:\ \begin{cases} x'' = \alpha x' + \beta y' \\ y'' = \gamma x' + \delta y' \end{cases}.$$

Cayley 考虑先执行 T_1 后执行 T_2 而得的变换

$$T_2T_1:\ \begin{cases} x'' = (a\alpha + c\beta)x + (b\alpha + d\beta)y \\ y'' = (a\gamma + c\delta)x + (b\gamma + d\delta)y \end{cases}$$

在研究如此复合变换的表达方式的过程中,他导致定义矩阵的乘法,复合变换 T_2T_1 的系数矩阵是 T_2 的矩阵乘以 T_1 的矩阵之积.他继续研究这种复合的代数——矩阵代数——包括矩阵的逆矩阵.用单个符号 A 来表示变换的矩阵是这种新代数的本质记号.矩阵代数和行列式之间的一条链结很快得以建立,即为基本结果:

$$\det(AB)=(\det A)(\det B).$$

Cayley 相信矩阵代数将会得以发展,以至夺去行列式理论的光彩.他写道:"在我看来,有许多事情说明这一矩阵理论将优先于行列式理论."

射向这一讨论的一束美妙侧光来自同一时期的另一位卓越的英国数学家 C. Babbage,他建造了第一台现代计算机,抽象出计算的构成方法以及其代数结构和记号,几乎涉及那个时期数学中同样普遍的智力进展的所有部分.

数学家们还试图发展向量代数,但是没有找到关于两个向量乘积的自然的定义.最早的向量代数,含有非交换的向量乘积,是由 H. Grassmann 于 1844 年提出的.其后,Grassmann 引进了由一个列向量乘以一个行向量形成的所谓简单矩阵.

矩阵和线性变换保持着紧密的联系,而且从进入 20 世纪后的观点看,正好是所形成的线性变换一般理论的有限维子情形.矩阵还被视为一种非常有用的记号,然而开头热衷一阵之后便很少就其进行研究.更得以关注的是向量,它是物理以及许多数学分支的基本数学元素.向量空间的现代定义是 Peano 在 1888 年引入的.不久,以函数乃至线性变换为元素的抽象向量空间随之建立.

回顾 20 世纪 40 年代的一个转折是可以令人得以教益的.其间,线性代数一度被认定作为研究课题已寿终正寝,从而应埋葬于教科书之中;然而曾几何时,响应高速计算机问世提供的机遇,利用标准矩阵运算的非常快的算法便相继涌现;矩阵数值分析得以强调,矩阵研究进入了出头之日.Von Neumann 和 Goldstein 在 1947 年提出分

析舍入误差中的条件数.A.Turing——与 Von Neumann 齐名的研制存储程序计算机的另一位天才——在 1948 年给出了矩阵的 LU 分解.QR 分解的实用是在十年之后实现的.

本书的目的和特点

今天,矩阵在各个学术领域和重要应用课题中已经起着不可替代的作用,计算机在数值计算方面的使用中,矩阵计算占据着大部分时间,因此矩阵的基本概念、理论和方法对于称职的和培育新的高素质科学技术人才来说是必备的非常重要而基础的知识.

迄今,已有许许多多矩阵著作.然而有专门术语或人名命名的矩阵原本已经很多,新的命名的矩阵又在增加着,它们的论述分布在众多著作之中,有时为了寻找某种矩阵的定义和性质并不容易.

本书取名"特殊矩阵",目的是提供一部好的尽可能全面介绍各类命名矩阵的论著,使它无论在学术上还是在应用上有其自身的价值和起着独特的作用.目前,尚未见到国内外有同样名称和目的的著作.

本书有如下特点:

(1) 取材丰富.涵盖基本的和常用的有命名的矩阵,并且包含到目前为止的绝大多数(包罗一切似不可能)其它有命名的矩阵,能反映有关的最新进展.

(2) 理论严谨.择优论述和推证方法.突出重点,对于重要的矩阵类型加以重点的全面的论述.

(3) 贯穿应用.一方面,对大多数矩阵类型提供一定的应用背景;另一方面,从应用课题中引出某些矩阵类型,并讨论其具体的应用.

(4) 结构合理.既有系统性,从基础知识出发,循序渐进,适合全面阅读;又具可分性,便于划分层次,选读部分内容.

(5) 灵活实用.索引详尽,采用中英文合一(英文只给出有命名的矩阵术语),查找方便,各取所需,也是一本必要的备用工具书.

(6) 深入浅出.分析清新,富启发性,语言精炼.阅读本书只需具备微积分和线性代数的基本知识.

本书兼理论专著、工具书、大学有关专业教材或参考书于一身.读者对象主要为数学尤其是应用数学和数值数学、科学、工程技术以及经济科学等工作者、大学教师、本科生和研究生.

本书的内容结构

所谓"特殊矩阵",如前面所说,亦即众多的各类命名矩阵,总归有其各自的独特刻画(定义).从大的方面来说,它们大体上可以划分成两部分:一部分是通过含有不易直观识别的性质来刻画的,书中称之为**特性矩阵**或**性质矩阵**,例如正规矩阵;另一部分则是通过容易直观识别的模式来刻画的,书中称之为**特型矩阵**,例如对称矩阵.划分成这样两部分,为有分有合地有机安排本书的章节带来了一定的方便.

全书共分 10 章.

第 1 章是基础知识,包括线性空间、对偶性、线性映射、矩阵、行列式、谱论、Euclid 空间、赋范线性空间和凸性的基本概念.这一章从现代数学的观点阐述了线性代数的基本理论,为阅读后面各章提供了一个很好的基础.然而就其本身而言,因是一个完整的整体,也可以作为独立内容来阅读.

第 2 章引入若干基本术语和记号、基本矩阵,以及不可约矩阵、对角优势矩阵、酉矩阵、正规矩阵、病态矩阵和 Vandermonde 矩阵等基本性质矩阵.

第 3 章包括从二次型引出自伴矩阵及惯性律、自伴矩阵的谱分解及其特征值的变分性质、正定矩阵和半正定矩阵、稳定矩阵、还

有斜自伴矩阵等.

第 4 章讨论正矩阵、非负矩阵、循环矩阵和素矩阵、随机矩阵、M-矩阵、以及 H-矩阵等.

第 5 章提供 Jordan 标准形和相似变换、友矩阵和 Frobenius 矩阵、Schur 标准形和奇异值分解、Householder 变换和 Hessenberg 矩阵、Givens 变换和 QR 分解、以及 Gauss 变换和 LU 分解.所谓"标准型"和"变换"均系某类特殊矩阵.

第 6 章汇集带状矩阵、轮换矩阵、Toeplitz 矩阵、Hankel 矩阵、其它条纹矩阵、以及中心对称矩阵、同伴矩阵和结式矩阵等一批特型矩阵.

第 7 章致力 Kronecker 积和 Hadamard 积等特殊积矩阵,以及各种广义逆矩阵.

第 8 章专注矩阵迭代 Jacobi、Gauss-Seidel、SOR、SSOR、AOR 和 SAOR、以及 A DI 诸方法的矩阵分裂和迭代矩阵.

第 9 章论述多项式、非多项式和 Hadamard 三类矩阵函数,以及一般的函数矩阵和作为特殊情形的 λ-矩阵与有理矩阵.

第 10 章作为最后一章,综合介绍了矩阵的有向图和性质 A 矩阵、相合矩阵、辛矩阵、整数矩阵、奇偶校验矩阵、对合矩阵、区间矩阵、自反矩阵等众多矩阵,以及关于素矩阵、复对称矩阵和自伴矩阵等矩阵的进一步的性质.

末 结 束 语

写这本书是作者因多年科研和教学而深感其必须和可能的由来已久的心愿,但只是两年多前才有条件付诸实现,而书稿得以完成几乎延续了整整两年时间.

作者追求和渴望的是一部完美而好用的论著.然而,达到如此境地是一个"极限"过程.一方面,关于矩阵的文献资料不计其数,只能

参看其一小部分比较重要的而已，况且新文献新命名的矩阵还会不断出现，因此实际上只可能确保理论上和应用上重要而常用的矩阵在内的尽量多的特殊矩阵，而对于这些矩阵又只可能择其基本和主要的结论.另一方面，在内容结构和论述上，作者已经精心尽力，却深知依然有待精雕细刻.

因此，作者的实际期望是在读者面前奉献一件已然加工得比较光滑成型的好"坯子"，而且能得到大量指正和指点的最好回报，以便在有机会再版时进一步作深加工.从这一意义上说，本书尚未完结.

必须指出，梁国珍教授自始至终参与本书编著工作，并且极其认真和精益求精地承担了全书的校阅任务，是本书编写及保证其质量的一位难得的无名作者，功不可没.

最后，作者趁此真诚而郑重地感谢清华大学出版社对本书出版的热忱支持和帮助.

作　者

2000年夏秋之交

目　　录

1 基础知识

1.1 线性空间

1.1.1 定义 X 称为 K 上的**线性空间**,如果

(1) X 是非空集合,其元素称为**向量**,记作 x, y 等.

(2) K 是**数域**,即为某些复数之集,包括 0 和 1,而且对四则运算封闭——K 中任意两个数的和、差、积、商(除数不为零)仍属于 K, K 中的数记作 α, β 等.

(3) 规定了一个从积集 $X \times X$ 到 X 的映射称为**向量加法**,记作 $x+y$,满足

1) 结合律: $x+(y+z)=(x+y)+z$.

2) 交换律: $x+y=y+x$.

3) X 中存在唯一**零元素**,记作 0:

$$x+0=x, \quad \forall x \in X.$$

4) 每一 $x \in X$,在 X 中存在唯一**负元素**,记作 $-x$:

$$x+(-x)=0.$$

(4) 规定了一个从积集 $K \times X$ 到 X 的映射称为**数乘法**(数乘以向量),记作 αx,满足

1) 单位律: $1x=x$.

2) 结合律: $\alpha(\beta x)=(\alpha\beta)x$.

3) 分配律: $\alpha(x+y)=\alpha x+\alpha y$.

$$(\alpha+\beta)x=\alpha x+\beta x.$$

例 设 K 是任一数域.

(1) 含 n 个分量且每个分量均属于 K 的列向量全体的集合

$$K^n \equiv \left\{(\alpha_1,\cdots,\alpha_n)^{\mathrm{T}} : \alpha_1,\cdots,\alpha_n \in K\right\}, \qquad (1.1)$$

按逐个分量定义的加法和数乘,是 K 上线性空间,仍记作 K^n.

特别,$\mathbb{R}^n$ 和$\mathbb{C}^n$ 是本书中最常用的线性空间.$\mathbb{R}$是实数直线,$\mathbb{C}$是复数全体.

(2) 开区间 (a,b) 上具有连续 n 阶导数的实值函数(或复值函数) $f:(a,b)\to\mathbb{R}$(或 $f:(a,b)\to\mathbb{C}$)全体的集合

$$C^n(a,b) \equiv \left\{f : f^{(n)} \text{ 在} \left(a,b\right) \text{上连续}\right\}, \qquad (1.2)$$

按通常函数的加法和数乘,是 $K=\mathbb{R}$(或 $K=\mathbb{C}$)上线性空间,仍记作 $C^n(a,b)$. (a,b) 可取 $\mathrm{R}\equiv(-\infty,+\infty)$.

特别

$C(a,b)\equiv C^0(a,b)$ 是 (a,b) 上连续函数的线性空间;

$C^\infty(a,b)$ 是 (a,b) 上无穷可微函数的线性空间.

类似地,对闭区间 $[a,b]$,有记号 $C^n[a,b]$.

(3) 系数属于 K 次数小于 n 的多项式全体的集合

$$P_{n-1} \equiv \left\{p : p(x) = \sum_{i=0}^{n-1} \alpha_i x^i, \alpha_0,\cdots,\alpha_{n-1} \in K, x \in K\right\}, \qquad (1.3)$$

按多项式的加法和数乘,是 K 上线性空间

1.1.2 定义 设 X 和 Y 是数域 K 上的两个线性空间,如果存在一一对应 $f:X\to Y$,使得

$$f(x'+x'') = f(x') + f(x''), \quad \forall x',x'' \in X,$$

$$f(\alpha x) = \alpha f\left(x\right), \qquad \forall x \in X, \alpha \in K,$$

则称 X 和 Y **同构**,称 f 为 X 和 Y 的**同构映射.**

同构的线性空间借助线性空间中能够进行的运算,可以不加区别.容易举出许多这样的例子,按非常不同方式提供的两个线性空间

是同构的.

例 1.1.1 的例(1)和例(3)中的 K^n 和 P_{n-1} 同构(证明留给练习).

1.1.3 定义 设 V 是数域 K 上线性空间 X 的非空子集，如果

$$\alpha x+\beta y\in V \quad \forall x,y\in V;\ \alpha,\beta\in K,$$

则 V 也是数域 K 上的一个线性空间(证明留给读者),称为 X 的**子空间.**

例 (1) $X=K^n$; $V=\left\{(0,\alpha_2,\cdots,\alpha_{n-1},0):\alpha_2,\cdots,\alpha_{n-1}\in K\right\}$.

(2) $X=C(-\infty,+\infty)$; $V=P_{n-1}, K=(-\infty,+\infty)$.见(1.3).

1.1.4 定理 设 V 和 W 是数域 K 上线性空间 X 的子空间,则交 $V\cap W$ 也是 X 的子空间.

证 因 $0\in V, 0\in W$,有 $0\in V\cap W$,故 $V\cap W$ 非空.另一方面,若 $x,y\in V\cap W$,则由 $x,y\in V$, $x,y\in W$,推出对于任意的 $\alpha,\beta\in K$ 有

$$\alpha x+\beta y\in V,\quad \alpha x+\beta y\in W,$$

故 $\alpha x+\beta y\in V\cap W$.因此 $V\cap W$ 是子空间. □

1.1.5 定义 设 V 和 W 是数域 K 上线性空间 X 的子集,集

$$\{x+y:x\in V, y\in W\}$$

称为 V 与 W 的**和**，记作 $V+W$.

1.1.6 定理 设 V 和 W 是数域 K 上线性空间 X 的子空间,则和 $V+W$ 也是一个子空间.

证 显然 $0\in V+W$,故 $V+W$ 非空.又若 $x,y\in V+W$,则存在 $x',y'\in V$, $x'',y''\in W$,使得 $x=x'+x'', y=y'+y''$.于是对于任意的 $\alpha,\beta\in K$,有

$$\alpha x' + \beta y' \in V, \quad \alpha x'' + \beta y'' \in W.$$

从而

$$\alpha x + \beta y = (\alpha x' + \beta y') + (\alpha x'' + \beta y'') \in V + W. \qquad \square$$

1.1.7 定义 设 $x^{(1)},\cdots,x^{(m)}$ 是数域 K 上线性空间 X 的 m 个向量，形如

$$\alpha_1 x^{(1)} + \cdots + \alpha_m x^{(m)}, \quad \alpha_1,\cdots,\alpha_m \in K$$

的向量称为 $x^{(1)},\cdots,x^{(m)}$ 的**线性组合**.

容易证明(留作练习),形如定义中的线性组合的全体构成 X 的一个子空间.

1.1.8 定义 由数域 K 上线性空间 X 中 m 个向量 $x^{(1)},\cdots,x^{(m)}$ 的所有线性组合构成的子空间,称为 $x^{(1)},\cdots,x^{(m)}$ 的**生成子空间**,记作 $\operatorname{span}\left\{x^{(1)},\cdots,x^{(m)}\right\}$,即

$$\operatorname{span}\left\{x^{(1)},\cdots,x^{(m)}\right\} = \left\{x \equiv \sum_{i=1}^{m} \alpha_i x^{(i)} : \alpha_1,\cdots,\alpha_m \in K\right\}. \tag{1.4}$$

1.1.9 定义 数域 K 上线性空间 X 中 m 个向量 $x^{(1)},\cdots,x^{(m)}$ 称为**线性相关**,如果它们之间存在一个如下形式的非平凡线性关系

$$\alpha_1 x^{(1)} + \cdots + \alpha_m x^{(m)} = 0,$$

其中 $\alpha_1,\cdots,\alpha_m \in K$ 不全为零.相反,如果向量 $x^{(1)},\cdots,x^{(m)}$ 非线性相关,则称它们为**线性无关**.

1.1.10 引理 设 $X = \operatorname{span}\left\{x^{(1)},\cdots x^{(n)}\right\}$,向量 $y^{(1)},\cdots,y^{(m)} \in X$ 线性无关,则 $m \le n$.

证 因 X 的每个向量可以写成 $x^{(1)},\cdots,x^{(n)}$ 的线性组合,又依线性无关的定义易知 $y^{(1)},\cdots,y^{(m)}$ 全为非零向量,故 $y^{(1)}$ 有表示式

$$y^{(1)} = \alpha_1 x^{(1)} + \cdots + \alpha_n x^{(n)},$$

且$\alpha_1, \cdots, \alpha_n$不全为零.不妨设$\alpha_1 \neq 0$,于是$x^{(1)}$可以表示成$y^{(1)}$与$x^{(2)}, \cdots, x^{(n)}$的线性组合,而且$\text{span}\{y^{(1)}, x^{(2)}, \cdots, x^{(n)}\} = X$.

假若$m \geq n$,容易推出,再重复上述过程$n-1$次便可以得到$\text{span}\{y^{(1)}, \cdots, y^{(n)}\} = X$.由此,如果$m > n$,那么立即得出与$y^{(1)}, \cdots, y^{(m)} \in X$线性无关相悖. □

1.1.11 定义 设一个由线性无关向量组成的有限集,生成线性空间X,则称该集为X的一组**基**.而且,有基的线性空间称为是**有限维的**.

1.1.12 定理 有限个向量$x^{(1)}, \cdots, x^{(n)}$生成的线性空间X必有基.

证 若$x^{(1)}, \cdots, x^{(n)}$线性相关,则它们之间存在非平凡关系,并由其可得$x^{(1)}, \cdots, x^{(n)}$中的某一向量$x^{(s)}$能表示成其余向量的线性组合.于是可以去掉$x^{(s)}$.重复这一步骤,直至余下的那些$x^{(i)}$彼此线性无关;它们仍然生成X,而且构成一组基. □

有限维线性空间有许许多多的基.

1.1.13 定理 设X是数域K上的有限维线性空间.则

(1) X所有基所包含的向量个数是相同的.这个数称为X的**维数**,记作$\dim X$.

(2) X同构于K^n,这里$n = \dim X$.由此推出,同一个数域上的任何两个维数相同的线性空间是同构的.

证(1) 从**1.1.10**和**1.1.11**直接推出.

(2)的证明留作练习. □

1.1.14 定理 有限维线性空间X中每一组线性无关向量$x^{(1)}, \cdots, x^{(s)}$可以扩充成X的一组基.

证 如果$x^{(1)}, \cdots, x^{(s)}$不生成X,那么存在某$x \in X$不能表示

成 $x^{(1)},\cdots,x^{(s)}$ 的线性组合.于是得扩充组 $x^{(1)},\cdots,x^{(s)},x^{(s+1)}\equiv x$. 重复这一步骤,直至扩充组生成 X.这一扩充过程在 $n-s$ 步内完成,$n=\dim X$. □

此定理表明一个有限维线性空间可以形成许多不同的基.

1.1.15 定理 设 X 是有限维线性空间,Y 是 X 的子空间.则

(1) Y 是有限维的.

(2) Y 有一个**补**——X 的另一子空间 Z,使得每个 $x\in X$ 可以唯一地分解成

$$x=y+z,\quad y\in Y,\quad z\in Z,$$

而且

$$\dim X=\dim Y+\dim Z. \tag{1.5}$$

证 在 Y 中可从其任一非零向量 $y^{(1)}$ 开始构造一组基,亦即得到生成 Y 的一个线性无关向量极大组 $y^{(1)},\cdots,y^{(s)}$. 依 **1.1.10**, $s\le\dim X$.再依 **1.1.14**,此极大组可以通过添加 $y^{(s+1)},\cdots,y^{(n)}$ 扩充成 X 的一组基.定义

$$Z\equiv\operatorname{span}\left\{y^{(s+1)},\cdots,y^{(n)}\right\}.$$

显然,Y 与 Z 相互为补,而且

$$\dim X=n=s+(n-s)=\dim Y+\dim Z. \qquad \square$$

比(1.5)更一般的一个结果是(证明留作练习):设 X 是有限维线性空间,V 与 W 是 X 的子空间,而且 $X=V+W$,则

$$\dim X=\dim V+\dim W-\dim(V\cap W). \tag{1.6}$$

1.1.16 定义 线性空间 X 称为其相互为补的两个子空间 Y 和 Z 的**直和**,记作 $X=Y\oplus Z$.

更一般地, X 称为其子空间 $Y_1,\cdots,Y_m$ 的**直和**,记作

$$X=Y_1\oplus\cdots\oplus Y_m,$$

如果每个 $x\in X$ 可以唯一地分解成

$$x = y^{(1)} + \cdots + y^{(m)}; \quad y^{(1)} \in Y_1, \quad \cdots, \quad y^{(m)} \in Y_m .$$

容易证明:若 X 是有限维线性空间且 $X = Y_1 \oplus \cdots \oplus Y_m$,则

$$\dim X = \sum_{i=1}^{m} \dim Y_i \tag{1.7}$$

1.1.17 定义 设 Y 是线性空间 X 的子空间. 如果 $x', x'' \in X$ 满足 $x' - x'' \in Y$,则称 x' 与 x'' 为**同余模** Y,记作 $x' \equiv x'' \bmod Y$.

容易证明:同余模是一种等价关系,即具有

(1) 对称性: $x' \equiv x'' \bmod Y \Rightarrow x'' \equiv x' \bmod Y$.

(2) 自反性: $x \equiv x \bmod Y$.

(3) 传递性: 成立

$$x' \equiv x'' \bmod Y, x'' \equiv x''' \bmod Y \Rightarrow x' \equiv x''' \bmod Y.$$

可以将 X 的元素分成同余类 $\bmod Y$.包含 $x \in X$ 的**同余类**记作 $\{x\}$,即

$$\{x\} \equiv \{z : z \in X, z - x \in Y\}. \tag{1.8}$$

1.1.18 定义 设 Y 是线性空间 X 的子空间. X 的同余类 $\bmod Y$ 所成之集在**加法**

$$\{x\} + \{y\} \equiv \{x + y\}$$

与**数乘法**

$$\alpha\{x\} \equiv \{\alpha x\}$$

的定义下构成一个线性空间(证明留作练习),称为 $X \bmod Y$ 的**商空间**,记作 $X(\bmod Y)$ 或 X / Y.

同余类加法定义是说:包含 x 的同余类与包含 y 的同余类之和就是包含 $x + y$ 的同余类.类似地可解释数乘法的定义.容易证明,这样的运算定义与同余类中代表元素的选取无关.

例 设线性空间 X 的组成元素是形如

$$(x_1,\cdots,x_n)$$

的具有 n 个分量的行向量，X 的子空间 Y 的元素形如

$$(0,0,x_3,\cdots,x_n),$$

即头两个分量为零的行向量.

此时,X 中任何两个向量 $x'=(x_1',\cdots,x_n')$ 与 $x''=(x_1'',\cdots,x_n'')$ 满足 $x'\equiv x''\bmod Y$ 的充分必要条件是 $x_1'=x_1'',x_2'=x_2''$(即只要头两个分量对应相等);于是,每个同余等价类可以用具有两个分量——等价类中所有向量的共同分量——的向量来表示.

此例表明,建立一个商空间 X/Y 相当于在 X 中的信息内抛弃其所包含的属于 Y 的那些成分.这种做法对于精简信息——忽略不必要的成分——来说是非常有意义的.

1.1.19 定理 设 Y 是线性空间 X 的子空间,则

$$\dim Y+\dim(X/Y)=\dim X. \tag{1.9}$$

特别,当 $\dim Y=\dim X$ 时,$Y=X$.

证 设 $y^{(1)},\cdots,y^{(s)}$ 是 Y 的一组基.依 **1.1.14**,可对其添加 $y^{(s+1)},\cdots,y^{(n)}$ 构成 X 的一组基.显然,$\{y^{(s+1)}\},\cdots,\{y^{(n)}\}$ 是 X/Y 的一组基. □

1.1.20 定义 设 X_1 与 X_2 是同一数域上的两个线性空间,形如

$$(x^{(1)},x^{(2)}),\quad x^{(1)}\in X_1,\quad x^{(2)}\in X_2$$

的有序对全体,在如下加法

$$(x^{(1)},x^{(2)})+(y^{(1)},y^{(2)})\equiv(x^{(1)}+y^{(1)},x^{(2)}+y^{(2)})$$

与数乘法

$$\alpha(x^{(1)},x^{(2)})\equiv(\alpha x^{(1)},\alpha x^{(2)})$$

的定义之下,构成一个线性空间(证明留作练习),称为 X_1 与 X_2 的**直和**,记作 $X_1\oplus X_2$.

容易证明(留作练习):

$$\dim(X_1 \oplus X_2) = \dim X_1 + \dim X_2 . \tag{1.10}$$

1.2 对偶性

1.2.1 定义 设 X 是数域 K 上的线性空间.定义在 X 上的纯量值函数 $l: X \to K$ 称为**线性**的,如果

$$l(x+y) = l(x) + l(y), \quad \forall x, y \in X , \tag{2.1}$$

而且

$$l(\alpha x) = \alpha l(x), \quad \forall x \in X, \alpha \in K . \tag{2.2}$$

注意,重复应用(2.1)与(2.2)这两条性质,可得

$$l(\alpha_1 x^{(1)} + \cdots + \alpha_n x^{(n)}) = \alpha_1 l(x^{(1)}) + \cdots + \alpha_n l(x^{(n)}),$$

$$\forall x^{(1)}, \cdots, x^{(n)} \in X; \alpha_1, \cdots, \alpha_n \in K . \tag{2.3}$$

1.2.2 定义 设 X 是数域 K 上的线性空间.定义在 X 上的线性函数全体,在逐点 $x \in X$ 定义的线性函数的加法

$$(l+m)(x) = l(x) + m(x) \ (l, m \text{ 是 } X \text{ 上的线性函数})$$

与数乘法

$$(\alpha l)(x) = \alpha l(x) \ (\alpha \in K; l \text{ 是 } X \text{ 上的线性函数})$$

之下,构成一个线性空间(证明留作练习),称为 X 的**对偶空间**,记作 X'.

例 (1) $X = C[0,1]$.定义 l_1,

$$l_1(f) \equiv \int_0^1 f(s) ds, \quad \forall f \in X .$$

任意取定 $s_0 \in [0,1]$,定义 l_2,

$$l_2(f) \equiv f(s_0), \quad f \in X .$$

任意取定$s_1,\cdots,s_n\in[0,1]$, $\alpha_1,\cdots,\alpha_n\in\mathbb{R}$,定义$l_3$,

$$l_3(f)\equiv\sum_{i=1}^{n}\alpha_i f(s_i),\quad \forall f\in X.$$

容易验证$l_1,l_2,l_3\in X'$.

(2) $X=C^\infty[0,1]$.任意取定$s_0\in[0,1]$,定义l,

$$l(f)\equiv\sum_{i=1}^{n}\alpha_i f^{(i)}(s_0),\quad \forall f\in X.$$

容易验证$l\in X'$.

1.2.3 定理 设X是数域K上的n维线性空间,$x^{(1)},\cdots,x^{(n)}$是X的一组基.则每个$x\in X$可唯一地表示为

$$x=\alpha_1 x^{(1)}+\cdots+\alpha_n x^{(n)},\quad \alpha_1,\cdots,\alpha_n\in K. \tag{2.4}$$

而且

(1) x的表示式(2.4)中的每个α_i是x的线性函数,记如此依赖关系为$\alpha_i=\alpha_i(x)$, $i=1,\cdots,n$.

(2) 对于任意给定的常数$c_1,\cdots,c_n\in K$,由

$$l(x)\equiv c_1\alpha_1(x)+\cdots+c_n\alpha_n(x) \tag{2.5}$$

定义的l是线性函数.

(3) 对于任意给定的非零向量$y\in X$,存在线性函数l,使得

$$l(y)\neq 0.$$

(4) 每个线性函数l可以唯一地写成(2.5)的形式.

证 (1)的证明留作练习.随之推出(2),因为线性函数的线性组合也是线性函数.

依 **1.1.14**,由非零向量$x^{(1)}\equiv y$出发可以扩充成X的一组基.由此从(1)推出(3).

设l是X上任一线性函数.将l作用于由(2.4)表示的x;利用线性函数的性质(2.3),立即看出l形如(2.5),其中

$$c_1 = l(x_1), \quad \cdots, \quad c_n = l(x_n). \qquad \square \quad (2.6)$$

1.2.4 定理 设 X 是数域 K 上的 n 维线性空间,则

$$\dim X' = \dim X.$$

证 关系式(2.4),(2.5)和(2.6)建立了线性函数 l 和 n 元有序数组 $(c_1, \cdots, c_n)$ 之间的一个一一对应;显然,如此一一对应是 X 的对偶空间 X' 和线性空间 K^n 之间的一个同构.另一方面,X 又和 K^n 同构.同构的线性空间是有相同的维数的. □

因为 X' 是线性空间,它也有自身的对偶空间 $X'' \equiv (X')'$.

1.2.5 定理 X'' 同构于 X.

证 对于固定的 $x \in X$,定义 X' 上的线性函数 ξ_x 如下:

$$\xi_x(l) \equiv l(x), \quad \forall l \in X'. \qquad (2.7)$$

可以直接验证,由(2.7)定义的 ξ_x 是 l 的线性函数.于是,对每个 $x \in X$,可按(2.7)定义一个线性函数;记如此线性函数所成之集为 Ξ.Ξ 是 X'' 的子空间.

首先证明由(2.7)确定的 X 和 Ξ 之间的对应是一一的,因而是一个同构.事实上,如果 $x, y \in X$,而且 $\xi_x = \xi_y$,即有

$$l(x) = \xi_x(l) = \xi_y(l) = l(y), \quad \forall l \in X',$$

那么

$$0 = l(x) - l(y) = l(x - y), \quad \forall l \in X',$$

于是,依 **1.2.3** 的(3),$x - y = 0$ 即 $x = y$.至此即可推出 X 和 Ξ 同构.依 **1.1.13**,

$$\dim X = \dim \Xi.$$

现在证明 Ξ 就是 X''.将 **1.2.4** 同时应用于 X 和 X',有

$$\dim X = \dim X' = \dim X'',$$

所以 $\dim \Xi = \dim X''$.依 **1.1.19**,$\Xi = X''$. □

这一定理表明 X 和 X' 之间的关系是完全对称的.为了强调这

一点,线性函数l在x的值有时记作

$$(l,x). \tag{2.8}$$

这是一个**双线性函数**,意即它对每个自变量当另一个被固定时是线性的;这是乘积的属性.因此,也称其为**点积**,记作

$$l\cdot x \quad 或 \quad lx. \tag{2.9}$$

1.2.6 定义 设Y是线性空间X的子空间,在Y上等于零的线性函数的集合,称为子空间Y的**零化子**,记作$Y^{\perp}$,即

$$Y^{\perp}\equiv\{l\in X':l(y)=0,\forall y\in Y\}. \tag{2.10}$$

1.2.7 定理 设Y是线性空间X的子空间,则

(1) $Y^{\perp}$是X'的子空间.

(2) $\dim Y+\dim Y^{\perp}=\dim X$.

证 (1)的证明留作练习.

考虑$Y^{\perp}$和$(X/Y)'$的一个自然同构.对于取定的$l\in Y^{\perp}$,定义$L\in(X/Y)'$如下:

$$L\{x\}=l(x), \quad \forall\{x\}\in X/Y. \tag{2.11}$$

从(1.8)和(2.10)推出L的定义是明确无误的,也就是说,并不依赖代表同余类的元素x的选取.

反之,对于任一取定的$L\in(X/Y)'$,(2.11)定义了一个$l\in Y^{\perp}$.显然,l和L的对应是一一的,从而是一个同构.于是

$$\dim Y^{\perp}=\dim(X/Y)', \tag{2.12}$$

依**1.2.4**,$\dim(X/Y)'=\dim(X/Y)$.至此,从(1.9)推出(2). □

$Y^{\perp}$的维数记作$\operatorname{codim}Y$,称为X的子空间Y的**余维数**.

依上述定理,有

$$\operatorname{codim}Y+\dim Y=\dim X. \tag{2.13}$$

由于$Y^{\perp}$是X'的子空间,它的零化子是X''的子空间,记作$Y^{\perp\perp}\equiv(Y^{\perp})^{\perp}$.

1.2.8 定理 在 X'' 和 X 的同一化(2.7)下,对 X 的每个子空间 Y 成立

$$Y^{\perp\perp}=Y. \tag{2.14}$$

证 将 **1.2.7** 同时应用于 Y 和 $Y^{\perp}$,有

$$\dim Y+\dim Y^{\perp}=\dim X$$

和

$$\dim Y^{\perp}+\dim Y^{\perp\perp}=\dim X'.$$

由此并注意到 $\dim X'=\dim X$,可得

$$\dim Y^{\perp\perp}=\dim Y,$$

从而 $Y^{\perp\perp}$ 和 Y 同构.另一方面,利用(2.7),

$$\xi_y(l)=l(y)=0,\quad \forall y\in Y,\ l\in Y^{\perp},$$

表明每个 $y\in Y$ 对应一个 $\xi_y\in Y^{\perp\perp}$.因此,在视如此相对应的 y 和 ξ_y 为同一的意义下,成立(2.14). □

1.2.9 定理 设 $I\subset\mathbb{R}$ 是一个区间,$t_1,\cdots,t_n\in I$ 是 n 个不同的点.则存在 n 个数 $m_1,\cdots,m_n\in\mathbb{R}$,使得成立求积公式

$$\int_I p(t)dt=m_1p(t_1)+\cdots+m_np(t_n),\quad \forall p\in P_{n-1}, \tag{2.15}$$

其中 P_{n-1} 是次数小于 n 的实系数多项式的集合,见(1.3).

证 依 **1.1.2**, P_{n-1} 同构于 $\mathbb{R}^n$, $\dim P_{n-1}=n$.定义线性函数 $l_1,\cdots,l_n\in(P_{n-1})'$,

$$l_i(p)=p(t_i),\quad \forall p\in P_{n-1};\quad i=1,\cdots,n. \tag{2.16}$$

现在证明 $l_1,\cdots,l_n$ 线性无关.假定有线性关系式

$$c_1l_1+\cdots+c_nl_n=0, \tag{2.17}$$

依 $l_1,\cdots,l_n$ 的定义,(2.17)意味着

$$c_1p(t_1)+\cdots+c_np(t_n)=0,\quad \forall p\in P_{n-1}. \tag{2.18}$$

定义多项式 $q_1,\cdots,q_n\in P_{n-1}$,

$$q_k(t) \equiv (t-t_1)\cdots(t-t_{k-1})(t-t_{k+1})\cdots(t-t_n), \quad k=1,\cdots,n$$

在(2.18)中依次取 p 为 $q_1,\cdots,q_n$,即得

$$c_1 = \cdots = c_n = 0,$$

这说明线性关系式(2.17)必是平凡的,因此 $l_1,\cdots,l_n$ 线性无关.

依 **1.2.4**,

$$\dim(P_{n-1})' = \dim P_{n-1} = n,$$

故 $l_1,\cdots,l_n$ 是 $(P_{n-1})'$ 的一组基.于是,P_{n-1} 上任一线性函数 l 可以表示为 $l_1,\cdots,l_n$ 的线性组合.特别,区间 I 上的积分是 P_{n-1} 上的一个线性函数,因此存在 $m_1,\cdots,m_n$,使得

$$\int_I(\bullet)dt = m_1 l_1(\bullet)+\cdots+m_n l_n(\bullet),$$

由此及(2.16),便证明了存在形如(2.15)的求积公式. □

1.3 线性映射

1.3.1 定义 设 X 和 Y 是数域 K 上的两个线性空间.从 X 到 Y 的一个映射 $T: X \to Y$ 称为**线性**的,如果它满足**加性**

$$T(x+y) = T(x)+T(y), \quad \forall x,y \in X \tag{3.1}$$

和满足**齐次性**

$$T(\alpha x) = \alpha T(x), \quad \forall x \in X, \alpha \in K, \tag{3.2}$$

X 称为 T 的**定义域空间**, Y 称为 T 的**目标空间**;在 T 下 X 的**象**称为 T 的**值域**,记作 $\mathscr{R}_T$,即

$$\mathscr{R}_T \equiv \{T(x): x \in X\}, \tag{3.3}$$

$\mathscr{R}_T$ 是 Y 的线性子空间(证明留作练习). $x \in X$ 在 T 下的象 $T(x)$ 常写成乘的形式 Tx;这样,加性便成了分配律.

通常,线性映射也称为**线性算子**.

例 (1) 任何同构映射是线性映射.

(2) $X = Y = P_{n-1}$,见(1.3);$T = \dfrac{\mathrm{d}}{\mathrm{d}x}$.

(3) $X = Y = C^{\infty}(\mathbb{R})$;$T$ 是任何线性微分算子.

(4) $X = Y = C[a,b]$;T 是$[a,b]$上任何线性积分算子.

(5) $X = \mathbb{R}^n, Y = \mathbb{R}^m$. $y = Tx$ 定义为

$$y_i = \sum_{j=1}^{n} t_{ij} x_j, \quad i = 1,\cdots,m. \tag{3.4}$$

这里 $y = (y_1,\cdots,y_m)^{\mathrm{T}} \in Y, \quad x = (x_1,\cdots,x_n)^{\mathrm{T}} \in X$.

1.3.2 定义 线性映射 $T: X \to Y$ 的**零空间**(也称为**核**)是 X 中由 T 映成0的元素的集合,记作 $\mathscr{N}_T$,即

$$\mathscr{N}_T \equiv \{x : Tx = 0, x \in X\}, \tag{3.5}$$

而且它是 X 的线性子空间(证明留作练习).

1.3.3 定理 设 $T: X \to Y$ 是线性映射,则

$$\dim \mathscr{N}_T + \dim \mathscr{R}_T = \dim X. \tag{3.6}$$

证 T 作用于商空间 $X/\mathscr{N}_T$ 定义为

$$T\{x\} = Tx, \quad \forall \{x\} \in X/\mathscr{N}_T,$$

则 T 是 $X/\mathscr{N}_T$ 和 $\mathscr{R}_T$ 的一个同构映射,于是

$$\dim(X/\mathscr{N}_T) = \dim \mathscr{R}_T.$$

另一方面,依 **1.1.19**,

$$\dim(X/\mathscr{N}_T) = \dim X - \dim \mathscr{N}_T,$$

即得(3.6). □

1.3.4 推论 设 $T: X \to Y$ 是线性映射.

(1) 若 $\dim Y < \dim X$,则

$$Tx = 0, \quad 对某些 x \in X, x \neq 0.$$

特别,取 $X=\mathbb{R}^n, Y=\mathbb{R}^m$;$T$ 的定义如(3.4).可得如下结论:如果 $m<n$,那么线性方程组

$$\sum_{j=1}^{n} t_{ij}x_j = 0, \quad i=1,\cdots,m \tag{3.7}$$

有非平凡解,即解中至少有一个 $x_j \neq 0$.

(2) 若 $\dim Y = \dim X$,且满足 $Tx=0$ 的只有向量 $x=0$,则

$$\mathscr{R}_T = Y.$$

特别,取 $X=Y=\mathbb{R}^n$;T 由

$$\sum_{j=1}^{n} t_{ij}x_j = y_i, \quad i=1,\cdots,n \tag{3.8}$$

定义,可得如下结论:如果齐次方程组

$$\sum_{j=1}^{n} t_{ij}x_j = 0, \quad i=1,\cdots,n \tag{3.9}$$

只有平凡解 $x_1=\cdots=x_n=0$,那么对于任何取定的

$$(y_1,\cdots,y_n)^{\mathrm{T}} \in Y,$$

非齐次方程组(3.8)有唯一解.

证 (1) 由于

$$\dim \mathscr{R}_T \leq \dim Y < \dim X,$$

因此从 **1.3.3** 得知,$\dim \mathscr{N}_T > 0$,也就是说,$\mathscr{N}_T$ 必包含某些非零向量.

(2) 依假设,$\mathscr{N}_T=\{0\}$,所以 $\dim \mathscr{N}_T=0$,因此从 **1.3.3** 以及从(2)的假设有

$$\dim \mathscr{R}_T = \dim X = \dim Y.$$

于是,根据 **1.1.19**,$\mathscr{R}_T = Y$. □

1.3.5 应用

(1) 设 X 是次数小于 n 的复系数多项式全体构成的空间,$Y=\mathbb{C}^n$.

选取 n 个不同的复数 $s_1,\cdots,s_n$,定义映射 $T:X\to Y$ 为

$$Tp=\left(p(s_1),\cdots,p(s_n)\right)^{\mathrm{T}},\quad \forall p\in X,$$

则 $\mathscr{N}_T=\{0\}$.事实上,

$$Tp=0\Leftrightarrow p(s_1)=0,\cdots,p(s_n)=0,$$

也就是说,p 以 $s_1,\cdots,s_n$ 为零点.然而次数小于 n 的多项式 p 不可能有 n 个不同的零点,除非 $p\equiv 0$.因此,依 **1.3.4** 的(2),T 的值域是整个 $Y=\mathbb{C}^n$;这表明对于任意指定的 $(c_1,\cdots,c_n)^{\mathrm{T}}\in\mathbb{C}^n$,有唯一的 $p\in X$,使得

$$p(s_1)=c_1,\cdots,p(s_n)=c_n.$$

(2) 设 X 是次数小于 n 的实系数多项式全体构成的空间,$Y=\mathbb{R}^n$.

选取 n 个两两不相交的区间 $I_1,\cdots,I_n\subset\mathbb{R}$.对于任意的 $p\in X$,定义 $\overline{p}_j$ 为 p 在 I_j 上的平均值:

$$\overline{p}_j=\frac{1}{|I_j|}\int_{I_j}p(s)\mathrm{d}s,\quad |I_j|=I_j \text{ 的长度}. \tag{3.10}$$

定义映射 $T:X\to Y$ 为

$$Tp=(\overline{p}_1,\cdots,\overline{p}_n)^{\mathrm{T}},$$

则 $\mathscr{N}_T=\{0\}$.

事实上,若 $Tp=0$,即成立

$$\overline{p}_j=0,\quad j=1,\cdots,n,$$

则 p 在 I_j 内变号,从而至少有一零点,$j=1,\cdots,n$.因 $I_1,\cdots,I_n$ 两两不相交,p 至少有 n 个不同的零点,再注意到 p 是次数小于 n 的多项式,故 $p\equiv 0$.

因此,依 **1.3.4** 的(2),T 的值域是整个 $Y=\mathbb{R}^n$;这表明对于任意指定的 $(c_1,\cdots,c_n)^{\mathrm{T}}\in\mathbb{R}^n$,有唯一的 $p\in X$,使得

$$\overline{p}_1=c_1,\cdots,\overline{p}_n=c_n.$$

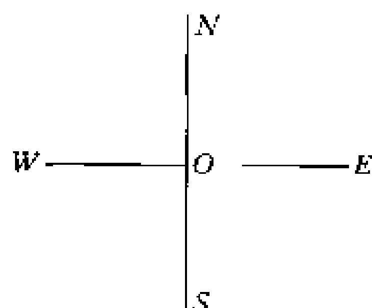
(3) 构造数值逼近 Laplace 方程在平面域 G 内的解,

$$\Delta u \equiv u_{xx} + u_{yy} = 0, \quad 在 G 内, \tag{3.11}$$

在 G 的边界上 u 是给定的.在 G 上引入方形网格,再用中心差分

$$u_{xx} \approx \frac{u_W - 2u_O + u_E}{h^2}, \quad u_{yy} \approx \frac{u_N - 2u_O + u_S}{h^2} \tag{3.12}$$

近似代替二阶偏导数,这里



而 h 是网格步长.将(3.12)代入(3.11)给出如下关系式:

$$u_O = \frac{u_W + u_N + u_E + u_S}{4}. \tag{3.13}$$

此方程表明域 G 内每个格点 O 处 u 的值 u_O 关系到 O 的四个相邻格点处 u 的值.在任何 O 的相邻格点处于 G 外的情形,取该点处 u 的值为 u 在最近边界点的边值.

方程(3.13)导致形如(3.8)的具有 n 个未知数 n 个方程的方程组;n 等于 G 内格点的个数.

现在证明相应的形如(3.7)的齐次方程组——相应于取边值为零——对所有格点只有零解 $u_O = 0$.

取定齐次方程组任一组解,并用 $u_{\max}$ 表示所有 G 内格点的 u_O 的最大值.假定在 G 内某格点 O 处达到 $u_{\max}$,则从(3.13)推出在 O 的四个相邻格点处全为 $u = u_{\max}$.重复如此论证,最终到达 G 外或边界的相邻格点.因为置所有这种点处的 u 为零,推出 $u_{\max} = 0$.类似可证 $u_{\min} = 0$.于是齐次方程组在所有格点的值

$u_O = 0$.

依 **1.3.4** 的(2),对于任何边界条件,导致的方程组有唯一解.

1.3.6 定义 设 T 和 S 是两个 $X \to Y$ 的线性映射;X 和 Y 是数域 K 上的线性空间.在**和** $T+S$:

$$(T+S)(x) \equiv T(x) + S(x), \quad \forall x \in X, \tag{3.14}$$

及**数乘积** αT:

$$(\alpha T) \equiv \alpha(Tx), \quad \forall x \in X, \alpha \in K \tag{3.15}$$

的规定下,$T+S$ 和 αT 仍是 $X \to Y$ 的线性映射,而且 $X \to Y$ 的线性映射的集合自身构成一个线性空间(证明留作练习).这个空间称为**线性映射空间**,记作 $\mathscr{L}(X,Y)$.

一般,设有映射(不一定是线性的)$T: X \to Y$ 和 $S: Y \to Z$,其中 X,Y 和 Z 是任意集合.T 和 S 的**复合**

$$S \circ T : X \to Z$$

定义为

$$(S \circ T)(x) \equiv S(T(x)), \quad \forall x \in X. \tag{3.16}$$

复合具有结合律:设有映射

$$T: X \to Y, S: Y \to Z, R: Z \to W,$$

则

$$R \circ (S \circ T) = (R \circ S) \circ T. \tag{3.17}$$

当 X、Y 和 Z 是线性空间,$T: X \to Y$ 和 $S: Y \to Z$ 是线性映射时,容易证明,$S \circ T$ 也是线性映射.而且,复合关于线性映射的加法具有分配律

$$(R+S) \circ T = R \circ T + S \circ T \tag{3.18}$$

和

$$S \circ (T+Q) = S \circ T + S \circ Q, \tag{3.19}$$

其中 $R: Y \to Z$ 和 $Q: X \to Y$ 也是线性映射.

由于有这种分配性质,连同一般成立的结合律,线性映射的复合常被表示为**乘法**的形式:

$$ST \equiv S \circ T. \tag{3.20}$$

注意,这种"乘法"一般是不可交换的.当 ST 有定义时,TS 经常是没有定义的,两者相等更是罕见.

1.3.7 定义 线性映射 $T: X \to Y$ 称为**可逆**的,如果它是一对一的和满的,即如果它是一个同构映射.T 的**逆**记作 $T^{-1}: Y \to X$,并且

$$T^{-1}(y) = x \text{ 当且仅当 } T(x) = y, \quad x \in X, y \in Y.$$

由定义推出(证明留作练习):

(1) 可逆线性映射的逆映射也是线性的.

(2) 若 T 和 S 是可逆的,而且 ST 有定义,则 ST 也是可逆的,而且有

$$(ST)^{-1} = T^{-1}S^{-1}. \tag{3.21}$$

1.3.8 定义 设 $T: X \to Y$ 是线性映射,X 和 Y 是数域 K 上的线性空间,$l: Y \to K$ 是线性函数,即 $l \in Y'$,则复合 $lT: X \to K$ 是线性函数,即 $lT \in X'$(证明留作练习).

这样,可以规定 $Y' \to X'$ 的一个映射,称为 T 的**转置**,记作 T',

$$T'l \equiv lT, \quad \forall l \in Y'. \tag{3.22}$$

更详细地

$$T'l(x) \equiv lT(x) = l(Tx), \quad \forall l \in Y', x \in X,$$

或者利用表示线性函数值的记号(2.8),

$$(T'l, x) \equiv (l, Tx), \quad \forall l \in Y', x \in X. \tag{3.23}$$

例 考虑 1.3.1 中的例(5): $X = \mathbb{R}^n, Y = \mathbb{R}^m, T: X \to Y$ 定义为 $y = Tx$,

$$y_i = \sum_{j=1}^{n} t_{ij} x_j, \quad i = 1, \cdots, m, \tag{3.24}$$

其中 $x = (x_1, \cdots, x_n)^{\mathrm{T}} \in X, \quad y = (y_1, \cdots, y_n)^{\mathrm{T}} \in Y$.

这时,对于 $l \in Y'$ 和 $f \in X'$ 分别有如下形式:

$$(l, y) = \sum_{i=1}^{m} l_i y_i \quad 和 \quad (f, x) = \sum_{j=1}^{n} f_j x_j .$$

可视 $l = (l_1, \cdots, l_m), \quad f = (f_1, \cdots, f_n)$,从而

$$(\mathbb{R}^n)' \equiv X' = \mathbb{R}^n, (\mathbb{R}^m)' \equiv Y' = \mathbb{R}^m .$$

由于

$$\begin{aligned}(l, Tx) = (l, y) = \sum_{i=1}^{m} l_i y_i &= \sum_{i=1}^{m} \sum_{j=1}^{n} l_i t_{ij} x_j \\ &= \sum_{j=1}^{n} \left(\sum_{i=1}^{m} l_i t_{ij} \right) x_j \equiv \sum_{j=1}^{n} f_j x_j = (f, x) \equiv (T'l, x),\end{aligned}$$

因此 $f = T'l$,

$$f_j = \sum_{i=1}^{m} l_i t_{ij}, \quad j = 1, \cdots, n. \tag{3.25}$$

转置有以下基本运算规律(证明留作练习):

$$(ST)' = T'S', \quad (T + R)' = T' + R', \quad (T^{-1})' = (T')^{-1} \tag{3.26}$$

而且,若在关系式(2.7)下视 X'' 恒同 X ,Y'' 恒同 Y ,则

$$T'' = T . \tag{3.27}$$

1.3.9 定理 线性映射 $T : X \to Y$ 的值域的零化子是其转置的零空间:

$$\mathscr{R}_T^{\perp} = \mathscr{N}_{T'} . \tag{3.28}$$

T 的值域是 T' 的零空间的零化子:

$$\mathscr{R}_T = \mathscr{N}_{T'}^{\perp} . \tag{3.29}$$

证 依 **1.2.6**,并注意到 $\mathscr{R}_T=\{y\equiv Tx : x\in X\}$ 以及表示式(3.23),有

$$\begin{aligned}\mathscr{R}_T^{\perp}&=\{l\in Y' : (l,y)=0,\forall y\in\mathscr{R}_T\}\\&=\{l\in Y' : (T'l,x)\equiv(l,Tx)=0,\forall x\in X\}.\end{aligned}\tag{3.30}$$

因此,对于 $\mathscr{R}_T^{\perp}$ 中的 l 有 $T'l=0$;这表明 $\mathscr{R}_T^{\perp}\subset\mathscr{N}_{T'}$.

另一方面,设 $l\in\mathscr{N}_{T'}$,则 $T'l=0$;由(3.30)即知 $l\in\mathscr{R}_T^{\perp}$.所以又有 $\mathscr{R}_T^{\perp}\supset\mathscr{N}_{T'}$.于是,(3.28)成立.

现在取(3.28)两边的零化子,并依 **1.2.8**,即得(3.29). □

(3.29)是映射值域的一个非常有用的特性.

1.3.10 定理 设 $T : X\to Y$ 是线性映射,则

$$\dim\mathscr{R}_T=\dim\mathscr{R}_{T'}.\tag{3.31}$$

如果还有 $\dim X=\dim Y$,那么

$$\dim\mathscr{N}_T=\dim\mathscr{N}_{T'}.\tag{3.32}$$

证 依 **1.2.7**,有

$$\dim\mathscr{R}_T^{\perp}+\dim\mathscr{R}_T=\dim Y,$$

将 **1.3.3** 应用于 $T' : X'\to Y'$,有

$$\dim\mathscr{N}_{T'}+\dim\mathscr{R}_{T'}=\dim Y'.$$

再依 **1.2.4**, $\dim Y=\dim Y'$,以及依 **1.3.9**, $\mathscr{R}_T=\mathscr{N}_{T'}^{\perp}$,立即得出(3.31).

将 **1.3.3** 同时应用于 T 和 T',注意到 $\dim Y=\dim Y'$ 及假设 $\dim X=\dim Y$,并利用(3.31),便可推出(3.32). □

1.3.11 定义 考虑线性空间 X 到其自身的线性映射构成的线性空间 $\mathscr{L}(X,X)$,这是特别重要的一类映射.任何两个这种映射可以相加与相乘(即复合),还可以乘以数.因此,$\mathscr{L}(X,X)$ 是一个**代数**.它是结合代数,但不是交换代数.它有一个单位元素——**恒等映射** I,

$$Ix = x, \quad \forall x \in X. \tag{3.33}$$

还有一个**零映射**O，

$$Ox = 0, \quad \forall x \in X. \tag{3.34}$$

$\mathscr{L}(X,X)$ 包含**零因子**——映射对T和S,其积$ST = O$但两者都不为零映射O.例如,S是具有非平凡零空间$\mathscr{N}_S$的任一非零映射,T是值域$\mathscr{R}_T \subset \mathscr{N}_S$的任一非零映射.还有,存在映射$T \neq O$,但$T^2 \equiv TT = O$;比如取$T$为微分算子,而$X$是次数小于 2 的多项式构成的线性空间,每个小于 2 的多项式的二阶导数等于零.

任意取定$A \in \mathscr{L}(X,X)$,经加法与乘法可形成A的多项式

$$p(A) \equiv \alpha_k A^k + \alpha_{k-1} A^{k-1} + \cdots + \alpha_0 I,$$

A的多项式全体的集合形成$\mathscr{L}(X,X)$的一个子代数.这个子代数是交换的,它在谱论中起着重要的作用.

若$T \in \mathscr{L}(X,X)$可逆,则依 **1.3.7**,显然有

$$TT^{-1} = T^{-1}T = I. \tag{3.35}$$

$\mathscr{L}(X,X)$的**可逆**元素的集合关于乘法(即复合)构成一个**群**.这个群只依赖于X的维数n以及数域K,记之为$GL(n,K)$.

设$S \in \mathscr{L}(X,X)$可逆,对每个$M \in \mathscr{L}(X,X)$可构造一个元素$M_S \in \mathscr{L}(X,X)$:

$$M_S \equiv SMS^{-1}. \tag{3.36}$$

这种从M到M_S的对应称为**相似变换**;并称M_S**相似**于M.

1.3.12 定理　每个相似变换是$\mathscr{L}(X,X)$的一个自同构,映射数乘到数乘,和到和,积到积,即成立

$$(\alpha M)_S = \alpha M_S. \tag{3.37}$$

$$(M+L)_S = M_S + L_S. \tag{3.38}$$

$$(ML)_S = M_S L_S. \tag{3.39}$$

相似变换构成一个群.另外

$$(M_S)_T = M_{TS}. \tag{3.40}$$

证 (3.37)和(3.38)显然成立.依相似变换定义(3.36),并利用结合律,有

$$(ML)_S = S(ML)S^{-1} = SMS^{-1}SLS^{-1} = M_S L_S$$

和

$$(M_S)_T = T(SMS^{-1})T^{-1} = (TS)M(TS)^{-1} = M_{TS},$$

后者还利用了关系式(3.21). □

1.3.13 定义 $P \in \mathscr{L}(X,X)$ 称为**投影映射**或**投影算子**,如果它具有等幂性,即满足

$$P^2 = P. \tag{3.41}$$

这是线性空间 X 到自身的一类重要映射.

例 (1) $X = \mathbb{R}^n$, P 的定义为

$$Px \equiv (0,0,x_3,\cdots,x_n)^{\mathrm{T}}, \quad \forall x = (x_1,\cdots,x_n)^{\mathrm{T}} \in X.$$

这是说,P 的作用是置换 x 的头两个分量为零.

(2) $X = C[-1,1]$,对于每个 $f \in C[-1,1]$,定义 Pf 为 f 的**偶部分**

$$(Pf)(x) \equiv \frac{f(x)+f(-x)}{2}, \quad \forall x \in [-1,1].$$

容易证明如此定义的 P 是线性的,而且是投影映射.

1.3.14 定理 设 $P \in \mathscr{L}(X,X)$ 是投影算子,则

(1) $I-P$ 也是 X 到 X 的投影算子,称为关于投影算子 P 的**余投影算子**.

(2) $\mathscr{R}_{I-P} = \mathscr{N}_P$.

(3) $\mathscr{N}_{I-P} = \mathscr{R}_P$.

(4) $\mathscr{N}_P \oplus \mathscr{R}_P = X$.

证 (1) 直接检验等幂性,

$$(I-P)^2 = I - 2P + P^2 = I - 2P + P = I - P.$$

(2) 若 $y \in \mathscr{R}_{I-P}$,则存在 $x \in X$,$(I-P)x = y$.因此

$$Py = P(I-P)x = (P-P^2)x = 0,$$

从而 $y \in \mathscr{N}_P$.

反之,若 $Py = 0$,则 $(I-P)y = y$,于是 $y \in \mathscr{R}_{I-P}$.

(3) 证明与(2)相仿,留作练习.

(4) 对于任一 $x \in X$,可将其写成

$$x = u + v;\quad u = (I-P)x \in \mathscr{N}_P, v = Px \in \mathscr{R}_P,$$

因此

$$\mathscr{N}_P + \mathscr{R}_P = X.$$

剩下需证明 x 的上述表示式是唯一的.事实上,若还有表示式

$$x = u' + v';\quad u' \in \mathscr{N}_P, v' \in \mathscr{R}_P,$$

则

$$u - u' = v' - v \in \mathscr{N}_P \cap \mathscr{R}_P.$$

于是 $u - u' \in \mathscr{N}_P$,即有

$$P(u-u') = 0.$$

而且 $u - u' \in \mathscr{R}_P$,依(3), 又有

$$(I-P)(u-u') = 0,$$

立即推出 $u - u' = 0$,从而

$$u = u', v = v'. \qquad \square$$

1.3.15 定义 线性空间 X 到 X 的两个映射 S 和 T 的**换位子**是指

$$ST - TS,$$

X 到 X 的两个映射 S 和 T 称为**(可)交换**的,如果其换位子是零(映射).

1.4 矩阵

1.4.1 定理 每个线性映射$T\in\mathscr{L}(\mathbb{C}^n,\mathbb{C}^m)$的对应$y=Tx$可以写成如下形式:

$$y_i=\sum_{j=1}^{n}t_{ij}x_j,\quad i=1,\cdots,m,\tag{4.1}$$

其中$x=(x_1,\cdots,x_n)^{\mathrm{T}}\in\mathbb{C}^n$, $y=(y_1,\cdots,y_m)^{\mathrm{T}}\in\mathbb{C}^m$.

证 设$e_j\in\mathbb{C}^n$是第j个分量为1其余分量为0的单位列向量, $j=1,\cdots,n$. $\{e_1,\cdots,e_n\}$称为$\mathbb{C}^n$(也是$\mathbb{R}^n$)中的**标准基**.

于是每个向量$x=(x_1,\cdots,x_n)^{\mathrm{T}}\in\mathbb{C}^n$可表示为

$$x=\sum_{j=1}^{n}x_je_j.$$

因T是线性的,

$$y=Tx=\sum_{j=1}^{n}x_jTe_j.\tag{4.2}$$

由此,并令

$$t_{ij}=(Te_j)_i,\quad i=1,\cdots,m,\quad j=1,\cdots,n$$

即得(4.1).注意,$Te_j\in\mathbb{C}^m$,$(Te_j)_i$是其第i个分量. □

1.4.2 定义 将出现在(4.1)的系数排列成矩形阵列

$$\begin{bmatrix}t_{11}&t_{12}&\cdots&t_{1n}\\t_{21}&t_{22}&\cdots&t_{2n}\\\cdots&\cdots&\cdots&\cdots\\t_{m1}&t_{m2}&\cdots&t_{mn}\end{bmatrix}\tag{4.3}$$

如此阵列称为 $m\times n$ **矩阵**, m 是**行**数, n 是**列**数,每个 t_{ij} 是 T 的**元素**. 这是线性映射 $T\in\mathscr{L}(\mathbb{C}^n,\mathbb{C}^m)$对应的矩阵,直接用 T 记之.特别,当 $m=n$ 时,称 T 为**方阵**,并称 n 为 T 的**阶**.

依 **1.4.1**,每个线性映射 $T\in\mathscr{L}(\mathbb{C}^n,\mathbb{C}^m)$对应一个形如(4.3)的 $m\times n$ 矩阵 T;反之,每个形如(4.3)的 $m\times n$ 矩阵 T 通过关系式(4.1)对应一个线性映射 $T\in\mathscr{L}(\mathbb{C}^n,\mathbb{C}^m)$.这就是说,线性映射集合 $\mathscr{L}(\mathbb{C}^n,\mathbb{C}^m)$和 $m\times n$ 矩阵全体之间存在一一对应.注意,在这种意义下,将把矩阵与线性映射**同一化**,许多概念和性质都可以用矩阵与线性映射两种语言来叙述.

t_{ij} 是 T 的在第 i 行和第 j 列交点处的元素,称为 T 的 (i,j)**-元素**,有时记之为 $(T)_{ij}$,即

$$(T)_{ij}\equiv t_{ij}. \tag{4.4}$$

通常,主要讨论所有元素 t_{ij} 为实数或为复数的矩阵.记号

$$T\in\mathbb{R}^{m\times n}$$

表示 T 的元素为实数,称 T 为 $m\times n$ **实矩阵**;类似地,记号

$$T\in\mathbb{C}^{m\times n}$$

表示 T 的元素为复数,称 T 为 $m\times n$ **复矩阵**.

T 是形如(4.3)的矩阵,有时将其视作**列向量的行**,或**行向量的列**:

$$T=\begin{bmatrix}c^{(1)} & \cdots & c^{(n)}\end{bmatrix}=\begin{bmatrix}r^{(1)}\\ \vdots\\ r^{(m)}\end{bmatrix}, \tag{4.5}$$

其中列向量

$$c^{(j)}=\begin{pmatrix}t_{1j}\\ \vdots\\ t_{mj}\end{pmatrix},\quad j=1,\cdots,n$$

和行向量

$$r^{(i)}=(t_{i1},\cdots,t_{in}),\quad i=1,\cdots,m.$$

而且也常用**缩写记号**

$$T=[t_{ij}].$$

根据(4.2)和(4.5),

$$Te_j=c^{(j)},\quad 1\le j\le n \tag{4.6}$$

表明列向量$e_j\in\mathbb{C}^n$的象是列向量$c^{(j)}\in\mathbb{C}^m$.因此,为了一致起见,一般约定对于任何正整数n,$\mathbb{C}^n$和$\mathbb{R}^n$中的任一向量x都是列向量,亦即

$$x=\begin{pmatrix}x_1\\ \vdots\\ x_n\end{pmatrix}\equiv(x_1,\cdots,x_n)^{\mathrm{T}}$$

两个$m\times n$矩阵$S=[s_{ij}]$与$T=[t_{ij}]$称为**相等**,如果

$$s_{ij}=t_{ij},\quad i=1,\cdots,m,\quad j=1,\cdots,n.$$

所有元素为零的矩阵称为**零矩阵**,记作 0.

容易证明,若$S,T\in\mathscr{L}(\mathbb{C}^n,\mathbb{C}^m)$,则

$$(S+T)_{ij}=(S)_{ij}+(T)_{ij}. \tag{4.7}$$

这样,自然地引出**矩阵加法**的定义:两个$m\times n$矩阵$S=[s_{ij}]$与$T=[t_{ij}]$之**和**记作$S+T$,

$$S+T\equiv[s_{ij}]+[t_{ij}]=[s_{ij}+t_{ij}]. \tag{4.8}$$

依 1.4.1,$\mathbb{C}^n \to \mathbb{C}^1$ 的线性映射 S 的对应矩阵是单行向量

$$s \equiv S = (s_1, \cdots, s_n).$$

$\mathbb{C}^1 \to \mathbb{C}^n$ 的线性映射 T 的对应矩阵是单列向量

$$t \equiv T = \begin{pmatrix} t_1 \\ \vdots \\ t_n \end{pmatrix}.$$

此时,积(复合)ST 是$\mathbb{C}^1 \to \mathbb{C}^1$ 的线性映射,容易推出,其对应矩阵是单个数(1×1 矩阵),记之为 st:

$$ST \equiv st = s_1t_1 + \cdots + s_nt_n. \tag{4.9}$$

从这个公式自然引出按行向量乘以列向量的**向量乘积**的定义.

应用定义(4.9)和(4.5)中的记号,可以通过矩阵作用于列向量给出公式(4.1)的一种紧凑写法:

$$y = Tx \equiv \begin{pmatrix} r^{(1)}x \\ \vdots \\ r^{(m)}x \end{pmatrix}. \tag{4.10}$$

下面说明应当怎样规定两个矩阵的乘法.

设 $T:\mathbb{C}^n \to \mathbb{C}^m$ 和 $S:\mathbb{C}^m \to \mathbb{C}^l$ 是线性映射,如前所述,对应的矩阵分别地仍用 T 和 S 来表示.同样地,对于复合映射 $ST:\mathbb{C}^n \to \mathbb{C}^l$,其对应的矩阵也同时用 ST 来表示.依(4.6)及复合映射定义,并应用(4.10),得到 ST 的第 j 列

$$(ST)e_j = S(Te_j) = Sc^{(j)} = \begin{pmatrix} s^{(1)}c^{(j)} \\ \vdots \\ s^{(l)}c^{(j)} \end{pmatrix},$$

其中 $s^{(k)}$ 是 S 的第 k 行.从线性映射角度解释如上关系式,便可自然地引出矩阵乘法的规则.

1.4.3 定义　矩阵乘法规则: 设 S 是 $l\times m$ 矩阵,T 是 $m\times n$ 矩阵. S 乘以 T 的**积**记作 ST,它是一个 $l\times n$ 矩阵,其 (k,j)-元素是 S 的第 k 行 $s^{(k)}$ 与 T 的第 j 列 $t^{(j)}$ 之积

$$(ST)_{kj}=s_k t_j. \tag{4.11}$$

这里

$$S\equiv\begin{bmatrix}s^{(1)}\\ \vdots\\ s^{(l)}\end{bmatrix},\quad T\equiv\begin{bmatrix}t^{(1)} & \cdots & t^{(n)}\end{bmatrix}.$$

注意,矩阵 S 乘以矩阵 T,必须满足 S 的列数等于 T 的行数,这种情形,称 S 和 T 为**可相乘矩阵**(conformable matrices).矩阵乘法没有交换律,即使是两个方阵相乘,一般也是不能交换的.

因为线性映射的复合有结合律,矩阵乘法也有结合律.

此前已约定,$\mathbb{C}^n$ 是具有 n 个实分量的列向量全体构成的空间.现在将$\mathbb{C}^n$ 的对偶空间$(\mathbb{C}^n)'$ 视同于具有 n 个实分量的行向量全体构成的空间.

$l\in(\mathbb{C}^n)'$ 作用于 $x\in\mathbb{C}^n$,依(2.8)用括号记作

$$(l,x)\equiv lx. \tag{4.12}$$

设　l,T,x 是如下线性映射:

$$l:\mathbb{C}^m\to\mathbb{C},\ \ T:\mathbb{C}^n\to\mathbb{C}^m,\ \ x:\mathbb{C}\to\mathbb{C}^n.$$

根据结合律,

$$(lT)x=l(Tx). \tag{4.13}$$

由于 $l\in(\mathbb{C}^m)'$, $lT\in(\mathbb{C}^n)'$,利用记号(4.12)可将(4.13)写成

$$(lT,x)=(l,Tx). \tag{4.14}$$

回顾线性映射 T 的转置 T' 的定义 **1.3.8**,有

$$(T'l,x)=(l,Tx). \tag{4.15}$$

比较(4.14)和(4.15)可以得出:T'对应的矩阵应当是T对应的矩阵经行变成列、列变成行后所得的矩阵.由此自然地引出矩阵转置的定义. 矩阵T的**转置矩阵**记作T^{T},其(i,j)-元素即为T的(j,i)-元素:

$$\left(T^{\mathrm{T}}\right)_{ij}=(T)_{ji} \tag{4.16}$$

1.4.4 定义 设矩阵T形如(4.5),将(4.6)代入(4.2),得

$$y=Tx=x_1c^{(1)}+\cdots+x_nc^{(n)},$$

这表明矩阵T各列的所有线性组合组成T的值域.此值域是一个线性空间,它的维数在老式教科书中称为T的**列秩**.类似地可以定义T的**行秩**.(4.16)说明T的行秩是T^{T}的值域的维数.

依 **1.3.10**,

$$\dim\mathscr{R}_T=\dim\mathscr{R}_{T^T}$$

即可推得矩阵T的列秩和行秩是相等的,称其为矩阵T的**秩**.记作$\operatorname{rank}T$.

现在考虑怎样用矩阵表示任意的映射.

设$T:X\to Y$是任一映射,X同构于$\mathbb{C}^n$,$n=\dim X$,Y同构于$\mathbb{C}^m$,$m=\dim Y$.

仿 **1.4.1**, 在X中选取一组基$u^{(1)},\cdots,u^{(n)}$后,便可建立线性映射$\mathscr{L}(X,Y)$与$m\times n$矩阵之间的对应关系.T对应的矩阵仍记作T.

同构映射可以这样来产生:对于X中选定的基$u^{(1)},\cdots,u^{(n)}$,然后按照

$$u^{(i)}\leftrightarrow e_i,\quad i=1,\cdots,n$$

便可确定一个同构映射,记其为

$$B:X\to\mathbb{C}^n. \tag{4.17}$$

类似地,有同构映射

$$C:Y\to\mathbb{C}^m. \tag{4.18}$$

显然,因存在许多基而存在着许多同构映射.可以利用任何这样的同构映射表示T为$\mathbb{C}^n\to\mathbb{C}^m$的映射;记对应于同构映射$B$和$C$的$T$的$\mathbb{C}^n\to\mathbb{C}^m$的映射为$M$,容易推出

$$M=CTB^{-1}. \tag{4.19}$$

当T是$X\to X$的映射时,取$C=B$,(4.19)成为

$$M=BTB^{-1}. \tag{4.20}$$

此时,若$C:X\to\mathbb{C}^n$是另一个同构映射,并记对应于C的T的$\mathbb{C}^n\to\mathbb{C}^n$的映射为$N$,则利用结合律及(4.20),有

$$N\equiv CTC^{-1}=CB^{-1}BTB^{-1}BC^{-1}=SMS^{-1}, \tag{4.21}$$

其中$S\equiv CB^{-1}$.因B和C同为$X\to\mathbb{C}^n$的同构映射,故S是$\mathbb{C}^n\to\mathbb{C}^n$的映射而且是可逆的.(4.21)给出了关于$T:X\to X$的$\mathbb{C}^n\to\mathbb{C}^n$的两种映射表示$M$和$N$之间的关系.

以上论述,一旦引入逆矩阵概念,便可直接改换为矩阵说法.

1.4.5 定义 设T是$n\times n$矩阵.如果T所对应的线性映射是可逆的,并将其逆映射对应的矩阵记作T^{-1},则称T为**可逆矩阵**,称T^{-1}为T的**逆矩阵**.可逆矩阵也称为**非奇异矩阵**或**非退化矩阵**.相应地,不可逆矩阵称为**奇异矩阵**或**退化矩阵**.

显然,对于可逆矩阵T,成立形如(3.35)的基本关系式,即

$$TT^{-1}=T^{-1}T=I. \tag{4.22}$$

通常,多用满足条件(4.22)直接定义T的逆矩阵T^{-1}.

存在着如同(4.21)的关系的方阵M和N称为是**相似**的.

前面分析表明:相似的矩阵描述同一个从某空间到其自身的映射.可以期望相似的矩阵有相同的内蕴性质.

1.4.6 定义 设 $A=[a_{ij}]$ 是 $n\times n$ 矩阵. A 的从左上角至右下角沿

$$i=j, \quad j=1,\cdots,n$$

的线段称为 A 的**主对角线**,位于主对角线上的元素

$$a_{ii}(i=1,\cdots,n)$$

称为 A 的**(主)对角元素**. A 的从右上角至左下角沿

$$(i,n-i+1), \quad i=1,\cdots,n$$

的线段称为 A 的**次对角线**,它垂直于主对角线.平行于主对角线沿

$$|i-j|=\text{常数}$$

的线段称为 A 的**对角线**.

$n\times n$ 矩阵 D 称为**对角矩阵**,如果

$$(D)_{ij}=0, \quad \forall i\neq j, \quad 1\leq i,j\leq n.$$

常记作

$$D=\operatorname{diag}(d_1,\cdots,d_n), \tag{4.23}$$

其中 $d_i=(D)_{ii}, i=1,\cdots,n$.

$n\times n$ 矩阵 I 称为**单位矩阵**,如果元素

$$(I)_{ij}=\begin{cases}0, & i\neq j,\\ 1, & i=j,\end{cases} \quad 1\leq i,j\leq n.$$

1.5 行列式与迹

1.5.1 引言 本节利用体积的直觉性质来定义方阵的行列式.

$\mathbb{R}^n$ 中的**单纯形**是具有 $n+1$ 个顶点的多面体,取其一个顶点于原点 O,其余顶点记作 $a^{(1)},\cdots,a^{(n)}$.顶点的顺序是紧要的,故而

称 $O,a^{(1)},\cdots,a^{(n)}$ 是**有序单纯形的顶点**.记如此有序单纯形为

$$S=(O,a^{(1)},\cdots,a^{(n)}),$$

S 称为**退化**的,如果它置于一个 $n-1$ 维子空间.

下面将涉及有序单纯形的两个几何属性:**定向**和**体积**.

一个非退化有序单纯形 S 具有两种定向中的一种:正或负.

S 称为**正定向**的,如果 S 可以连续地和非退化地变形成为**标准有序单纯形**

$$(O,e_1,\cdots,e_n),$$

其中 e_i 是$\mathbb{R}^n$ 的标准基中的第 i 个单位向量.这种变形意味着存在 n 个 t 的向量值函数

$$a^{(i)}(t),\quad 0\le t\le 1,\quad i=1,\cdots,n$$

使得

$$S(t)=(O,a^{(1)}(t),\cdots,a^{(n)}(t))$$

对所有 $t\in[0,1]$ 是非退化的,而且满足

$$a^{(i)}(0)=a^{(i)},\quad a^{(i)}(1)=e_i,\quad i=1,\cdots,n$$

相反,即 S 不是正定向的,称 S 为**负定向**的.

对于有序单纯形 S,定义 $O(S)$ 如下:当 S 非退化时,依其定向的正和负,$O(S)$ 分别等于 $+1$ 和 -1;当 S 退化时,$O(S)=0$.

单纯形 S 的**体积**的基本公式是

$$\operatorname{vol}(S)=\frac{1}{n}\operatorname{vol}_{n-1}(底)\times 高 \tag{5.1}$$

其中"底"是指 S 的任何 $(n-1)$ 维面,"高"是指从包含底的超平面到对顶点的距离.

一个更为有用的概念是**带符号体积**,记作 $\Sigma(S)$,定义为

$$\Sigma(S)=O(S)\mathrm{vol}(S). \tag{5.2}$$

因为 S 是通过它的顶点来描述的,$\Sigma(S)$ 是 $a^{(1)},\cdots,a^{(n)}$ 的函数.显然,当两个顶点相等时,S 是退化的,所以

$$\Sigma(S)=0,\quad 当存在 a^{(i)}=a^{(j)},i\neq j. \tag{5.3}$$

$\Sigma(S)$ 的另一个性质是:当所有 $a^{(i)}(i\neq j)$ 固定时,$\Sigma(S)$ 是 $a^{(j)}$ 的线性函数.为了证明这一结论,将(5.1)和(5.2)合并为

$$\Sigma(S)=\frac{1}{n}\mathrm{vol}_{n-1}(底)\times\alpha, \tag{5.4}$$

其中

$$\alpha=O(S)\times 高. \tag{5.5}$$

这里高是包含底的超平面到顶点 $a^{(j)}$ 的距离;α 称为**带符号距离**,因为 $O(S)$ 当 $a^{(j)}$ 处于底的一个侧面时取一个符号,当 $a^{(j)}$ 处于底的相反侧面时取相反符号.引进笛卡儿坐标轴,使得第一轴垂直于底,其余轴置于底平面之中.按照笛卡儿坐标的定义,向量 a 的第一坐标 $\alpha_1(a)$ 是 a 到其余轴生成的超平面的带符号距离.根据 1.2.3 的(1),$\alpha_1(a)$ 是 a 的线性函数.由此,从(5.4)和(5.5)便推出所要证明的结论.

下述经典公式把行列式和有序单纯形的带符号体积联系在一起:

$$\Sigma(S)=\frac{1}{n!}D(a^{(1)},\cdots,a^{(n)}), \tag{5.6}$$

其中记号 $D(a^{(1)},\cdots,a^{(n)})$ 是在稍后 1.5.7 中将予以正式定义的以 $a^{(1)},\cdots,a^{(n)}$ 为列的行列式的缩写.对于行列式,不是从一个抽象公式出发,而是从带符号体积的几何性质赋予它的性质来演绎它,这种研究途径归功于 E.Artin.

1.5.2 性质 由(5.6)确定的 $D(a^{(1)},\cdots,a^{(n)})$ 具有如下性质:

(1) 若存在 $a^{(i)}=a^{(j)},i\neq j$,则

$$D(a^{(1)},\cdots,a^{(n)})=0.$$

(2) D 是关于 $a^{(1)},\cdots,a^{(n)}$ 的**多重线性函数**,也就是说,当所有 $a^{(i)}(i\neq j)$ 固定时,D 是 $a^{(j)}$ 的线性函数.

(3) 规范化:

$$D(e_1,\cdots,e_n)=1, \tag{5.7}$$

其中 $e_1,\cdots,e_n$ 是$\mathbb{R}^n$ 的标准基.

(4) D 是关于 $a^{(1)},\cdots,a^{(n)}$ 的**交错函数**,也就是说,如果交换 $a^{(i)}$ 和 $a^{(j)}$,$i\neq j$,那么 D 的值改变因子(-1).

(5) 若 $a^{(1)},\cdots,a^{(n)}$ 线性相关,则

$$D(a^{(1)},\cdots,a^{(n)})=0$$

证 (1)和(2)可直接从 $\Sigma(S)$ 的相应性质推出.(1)也是(5)的特殊情形.

(3) 可从(5.6)、(5.5)、(5.4)和(5.1)递推或归纳,具体证明留作练习.

(4) 因为只交换了 $a^{(i)}$ 和 $a^{(j)}$,下面仅指示它们,令

$$a\equiv a^{(i)},b\equiv a^{(j)}.$$

使用性质(1)和(2),得

$$\begin{aligned}D(a,b)&=D(a,b)+D(a,a)=D(a,a+b)\\&=D(a,a+b)-D(a+b,a+b)\\&=-D(b,a+b)=-D(b,a)-D(b,b)\\&=-D(b,a)\end{aligned}$$

(5) 因 $a^{(1)},\cdots,a^{(n)}$ 线性相关,故它们中之一,不妨设 $a^{(1)}$,可表示为其余者的线性组合:

$$a^{(1)}=\alpha_2a^{(2)}+\cdots+\alpha_na^{(n)}.$$

先使用性质(2),再使用性质(1),即得

$$\begin{aligned}&D(a^{(1)},\cdots,a^{(n)})\\&=D(\alpha_2a^{(2)}+\cdots+\alpha_na^{(n)},a^{(2)},\cdots,a^{(n)})\end{aligned}$$

$$= \alpha_2 D(a^{(2)}, a^{(2)}, \cdots, a^{(n)}) + \cdots$$

$$+ \alpha_n D(a^{(n)}, a^{(2)}, \cdots, a^{(n)}) = 0 . \qquad \square$$

1.5.3 定义 n 个对象,比如数1,2,$\cdots$,n,到其自身的满映射 p 称为**置换**.

如同所有函数那样,置换可以复合.因为置换是满映射,必然是一对一的,所以还可以求逆.因此它们形成一个群;此类群,除 $n=2$ 外,是不可交换的.

记 $p_i \equiv p(i)$.借助如下形式的表格显示 p 的作用很是方便:

1	2	$\cdots$	n
p_1	p_2	$\cdots$	p_n

简记作 $\dfrac{12\cdots n}{p_1 p_2 \cdots p_n}$.

例 设

$$p = \frac{1234}{2413},$$

则有

$$p^2 = \frac{1234}{4321}, \quad p^3 = \frac{1234}{3142}, \quad p^4 = \frac{1234}{1234}$$

和

$$p^{-1} = \frac{1234}{3142}.$$

1.5.4 定义 设 $x_1, \cdots, x_n$ 是 n 个变量;它们的**判别式**是指

$$P(x_1, \cdots, x_n) \equiv \prod_{i<j} (x_i - x_j) . \tag{5.8}$$

设 p 是任一置换.显然,

$$P(p(x_1,\cdots,x_n))=\prod_{i<j}(x_{p_i}-x_{p_j})$$

就是 $P(x_1,\cdots,x_n)$ 或 $-P(x_1,\cdots,x_n)$.

1.5.5 定义 设 p 是任一置换. p 的**符号差** $\sigma(p)$ 定义为

$$P(p(x_1,\cdots,x_n))\equiv\sigma(p)P(x_1,\cdots,x_n). \tag{5.9}$$

符号差的性质(证明留作练习):

(1) $\sigma(p)=+1$ 或 -1.

(2) $\sigma(p_1\circ p_2)=\sigma(p_1)\sigma(p_2)$.

置换的一种特殊类型:一次交换.它针对任何指标对

$$j,k,\quad j\neq k,$$

其定义如下:

$$\begin{aligned}&p(i)=i, \quad && i\neq j,k,\\ &p(j)=k, && p(k)=j,\end{aligned} \tag{5.10}$$

这种置换称为**对换**.

对换有以下性质(证明留作练习):

(3) 对换 t 的符号差

$$\sigma(t)=-1.$$

(4) 每个置换 p 可以写成对换的复合:

$$p=t_k\circ\cdots\circ t_1.$$

如此,

$$\sigma(p)=(-1)^k.$$

这种分解是不唯一的,但因子个数 k 的奇偶性是唯一的.

例 置换

$$p=\frac{12345}{24513}$$

是对换

$$t_1=\frac{12345}{12543},\quad t_2=\frac{12345}{21345},\quad t_3=\frac{12345}{42315}$$

的复合, $p=t_3\circ t_2\circ t_1$.

1.5.6 定理 设

$$a^{(j)}\equiv\begin{pmatrix}a_{1j}\\ \vdots\\ a_{nj}\end{pmatrix},\quad j=1,\cdots,n, \tag{5.11}$$

则由(5.6)确定的 $D(a^{(1)},\cdots,a^{(n)})$ 成立如下基本公式:

$$D(a^{(1)},\cdots,a^{(n)})=\sum_p\sigma(p)a_{p_1 1}\cdots a_{p_n n}. \tag{5.12}$$

这里等式右边求和是对 $1,\cdots,n$ 的所有置换 p 而言的.

证 因为

$$a^{(j)}=a_{1j}e_1+\cdots+a_{nj}e_n,\quad j=1,\cdots,n,$$

其中 $e_1,\cdots,e_n$ 是 $\mathbf{R}^n$ 的标准基. 依 **1.5.2** 的(2),并利用 $a^{(1)}$ 关于 $e_1,\cdots,e_n$ 的线性组合,得

$$\begin{aligned}&D(a^{(1)},\cdots,a^{(n)})\\&=D(a_{11}e_1+\cdots+a_{n1}e_n,a^{(2)},\cdots,a^{(n)})\\&=a_{11}D(e_1,a^{(2)},\cdots,a^{(n)})+\cdots\\&\quad+a_{n1}D(e_n,a^{(2)},\cdots,a^{(n)}),\end{aligned}$$

$D(a^{(1)},\cdots,a^{(n)})$ 展成 n 项;接着再利用 $a^{(2)}$ 关于 $e_1,\cdots,e_n$ 的线性组合,得 n^2 项展式;重复这一过程共 n 次,可得

$$D(a^{(1)},\cdots,a^{(n)})=\sum_f a_{f_1 1}\cdots a_{f_n n}D(e_{f_1},\cdots,e_{f_n}), \tag{5.13}$$

其中求和是对映射$\{1,\cdots,n\}$到$\{1,\cdots,n\}$的所有函数f而言的.若某f不是置换(即非满映射),则存在$i,j,i\neq j$有$f_i=f_j$;对于这种情形,依**1.5.2**的(1),

$$D(e_{f_1},\cdots,e_{f_n})=0. \tag{5.14}$$

这说明在(5.13)中只须对作为置换的那些f求和.依 **1.5.5** 的(4)及(5.7),对任何置换p,

$$D(e_{p_1},\cdots,e_{p_n})=\sigma(p)D(e_1,\cdots,e_n)=\sigma(p). \tag{5.15}$$

将(5.14)和(5.15)代入(5.13)即得(5.12). □

(5.12)的意义在于:它是D的关于其自变量的分量的展开公式.

1.5.7 定义 设$A=\begin{bmatrix}a^{(1)} & \cdots & a^{(n)}\end{bmatrix}$是$n\times n$矩阵,$a^{(1)},\cdots,a^{(n)}$是其列向量.$A$的**行列式**,记作$\det A$,是指

$$\det A\equiv D(a^{(1)},\cdots,a^{(n)}), \tag{5.16}$$

其中D由公式(5.12)给定.

行列式成立**1.5.2**中对D已作了证明的性质(1)~(5).下面陆续建立其它重要性质.

1.5.8 定理 设A和B是$n\times n$矩阵.则

$$\det(BA)=\det A\det B. \tag{5.17}$$

证 由于$Ae_j=a^{(j)}$是A的第j列,因此BA的第j列为

$$(BA)e_j=B(Ae_j)=Ba^{(j)}.$$

依(5.16),

$$\det(BA)=D(Ba^{(1)},\cdots,Ba^{(n)}). \tag{5.18}$$

现假定$\det B\neq 0$,定义函数C如下:

$$C(a^{(1)},\cdots,a^{(n)})\equiv\frac{\det(BA)}{\det B}. \tag{5.19}$$

下面证明 C 成立 **1.5.2** 中 D 所具有的性质(1) ~ (3).

(1) 若 $a^{(i)} = a^{(j)}, i \neq j$,则 $Ba^{(i)} = Ba^{(j)}$;因 D 具有 **1.5.2** 的性质(1),从(5.18)和(5.19)推出 $C(a^{(1)}, \cdots, a^{(n)}) = 0$.

(2) 因为 $Ba^{(j)}$ 是 $a^{(j)}$ 的线性函数,而且 D 是多重线性函数,推出(5.18)的右端也是 $a^{(1)}, \cdots, a^{(n)}$ 的多重线性函数,所以 C 必是 $a^{(1)}, \cdots, a^{(n)}$ 的多重线性函数.

(3) 置 $a^{(j)} = e_j$, $j = 1, \cdots, n$.注意到 $Be_j = b^{(j)}$ 是 B 的第 j 列,得

$$C(e_1, \cdots, e_n) = \frac{D(Be_1, \cdots, Be_n)}{\det B} = \frac{D(b^{(1)}, \cdots, b^{(n)})}{\det B} = 1. \tag{5.20}$$

由 **1.5.6** 的证明中得知,满足 **1.5.2** 中的性质(1) ~ (3)的函数 C 相等于函数 D,因此

$$C(a^{(1)}, \cdots, a^{(n)}) = D(a^{(1)}, \cdots, a^{(n)}) = \det A.$$

将此式代入(5.19),证明了当 $\det B \neq 0$ 时成立(5.17).

当 $\det B = 0$ 时,定义

$$B(t) \equiv B + tI,$$

公式(5.12)表明 $D(B(t))$ 是 t 的 n 次多项式,t^n 的系数等于1.于是,使 $D(B(t))$ 等于零的 t 值不会多于 n 个;特别,因

$$D(B(0)) = D(B) = 0,$$

故对所有近乎零但不等于零的 t 值 $D(B(t)) \neq 0$.这样,对于如此 t 值,根据已有证明,$D(B(t)A) = \det A \det B(t)$,令 t 趋于零即得(5.17). □

行列式乘法性质(5.17)的几何意义是:线性映射 B 将每个单纯形映射成另一个单纯形,后者的体积是原始单纯形体积的 $|\det B|$ 倍.由此推出,任何开集在 B 下的象的体积是原始体积的 $|\det B|$ 倍.

1.5.9 引理 设 $A=[a_{ij}]$ 是第一列为 e_1 的 $n\times n$ 矩阵:

$$A=\left[\begin{array}{c|ccc}1 & * & * & * \\ \hline 0 & & & \\ \vdots & & A_{11} & \\ 0 & & & \end{array}\right] \tag{5.21}$$

这里 A_{11} 表示 A 的由元素 $a_{ij}, i>1, j>1$ 形成的 $(n-1)\times(n-1)$ 子矩阵.则

$$\det A=\det A_{11}. \tag{5.22}$$

证 由 **1.5.2** 的性质(1)和(2)推出:如果在矩阵中将其一列的若干倍加到另一列,那么改变后的矩阵的行列式和原始矩阵的行列式保持相等.从(5.21)看出,在 A 中,依此加第一列的适当倍数至其它各列,可将其第1行的后 $n-1$ 个元素消成零.因此

$$\det A=\det\left[\begin{array}{c|ccc}1 & 0 & 0 & 0 \\ \hline 0 & & & \\ \vdots & & A_{11} & \\ 0 & & & \end{array}\right]. \tag{5.23}$$

引进函数 C,

$$C(A_{11})\equiv\det\begin{bmatrix}1 & 0 \\ 0 & A_{11}\end{bmatrix}.$$

显然,C 满足 **1.5.2** 的性质(1) ~ (3).所以

$$C(A_{11})=\det A_{11},$$

由此及(5.23)即得(5.22). □

1.5.10 推论 设 A 是第 j 列为 e_i 的 $n\times n$ 矩阵,A_{ij} 是在 A 中删去第 i 行和第 j 列后所得的 $(n-1)\times(n-1)$ 子矩阵.则

$$\det A=(-1)^{i+j}\det A_{ij}. \tag{5.24}$$

证明留作练习.

一般,设 $A=[a_{ij}]$ 是 $n\times n$ 矩阵, A_{ij} 是在 A 中删去第 i 行和第 j 列后所得的 $(n-1)\times(n-1)$ 子矩阵.则称

$$M_{ij}\equiv\det A_{ij}$$

为 A 的**第 (ij) 子式**或关于**元素 a_{ij} 的子式**.而

$$(-1)^{i+j}M_{ij}\equiv(-1)^{i+j}\det A_{ij}$$

称为关于**元素 a_{ij} 的余子式**.

1.5.11 定理 设 $A=[a_{ij}]$ 是 $n\times n$ 矩阵, j 是1与 n 之间的某一指数.则

$$\det A=\sum_{i=1}^{n}(-1)^{i+j}a_{ij}\det A_{ij}\ , \qquad (5.25)$$

此关系式称为**行列式依列的 Laplace 展开式**.

证 为了记号简单,取 $j=1$.将 A 的第1列 $a^{(1)}$ 写成标准基的线性组合:

$$a^{(1)}=a_{11}e_1+\cdots+a_{n1}e_n,$$

利用多重线性,得

$$\begin{aligned}\det A&=D(a^{(1)},\cdots,a^{(n)})\\&=D(a_{11}e_1+\cdots+a_{n1}e_n,a^{(2)},\cdots,a^{(n)})\\&=a_{11}D(e_1,a^{(2)},\cdots,a^{(n)})\\&\quad+\cdots+a_{n1}D(e_n,a^{(2)},\cdots,a^{(n)}),\end{aligned}$$

再利用(5.24),即得(5.25)的 $j=1$ 的情形. □

1.5.12 定理 设 A 是 $n\times n$ 矩阵,

$$A=\begin{bmatrix}a_{11}&\cdots&a_{1n}\\\cdots&\cdots&\cdots\\a_{n1}&\cdots&a_{nn}\end{bmatrix}\equiv[a^{(1)}\quad\cdots\quad a^{(n)}],$$

$\det A \neq 0$,而且

$$b = \begin{pmatrix} b_1 \\ \vdots \\ b_n \end{pmatrix}, \qquad x = \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix},$$

则线性方程组

$$Ax = b \tag{5.26}$$

存在唯一解

$$x_j = \frac{\det A_j}{\det A}, \quad j = 1, \cdots, n. \tag{5.27}$$

这里 A_j 是用 b 替换 A 的第 j 列而得到的矩阵,即

$$A_j = \begin{bmatrix} a^{(1)} & \cdots & a^{(j-1)} & b & a^{(j+1)} & \cdots & a^{(n)} \end{bmatrix}.$$

(5.27)就是著名的 **Cramer 法则**.它可以进一步写成

$$x_j = \sum_{i=1}^{n} (-1)^{i+j} \frac{\det A_{ij}}{\det A} b_i, \quad j = 1, \cdots, n. \tag{5.28}$$

证 x 可以写成

$$x = \sum_{k=1}^{n} x_k e_k,$$

这里 $e_1, \cdots, e_n$ 是$\mathbb{R}^n$ 的标准基.(5.26)等价于

$$\sum_{k=1}^{n} x_k a^{(k)} = b.$$

利用行列式的多重线性,则

$$\det A_j = \det\left(a^{(1)}, \cdots, a^{(j-1)}, \sum_{k=1}^{n} x_k a^{(k)}, a^{(j+1)}, \cdots, a^{(n)}\right)$$

$$= \sum_{k=1}^{n} x_k \det(a^{(1)}, \cdots, a^{(j-1)}, a^{(k)}, a^{(j+1)}, \cdots, a^{(n)}).$$

因为行列式成立 **1.5.2** 的性质(1),上式右端只有第 j 项为非零项,于是有

$$\det A_j = x_j \det A.$$

此式当 $\det A \neq 0$ 时等价于(5.27).现在对于 $\det A_j$ 依其第 j 列应用 Laplace 展开式,得到

$$\det A_j = \sum_{i=1}^{n} (-1)^{i+j} b_i \det A_{ij}.$$

由此,利用(5.27)即得(5.28). □

1.5.13 定理 $n \times n$ 矩阵 A 可逆的充分必要条件是 $\det A \neq 0$.当 A 可逆时,逆矩阵 A^{-1} 有如下公式:

$$(A^{-1})_{ij} = (-1)^{i+j} \frac{\det A_{ji}}{\det A}, \quad i, j = 1, \cdots, n. \qquad (5.29)$$

证 设 $\det A \neq 0$,则(5.29)有意义.将 A^{-1} 作用于向量 b ,即

$$(A^{-1}b)_i = \sum_{j=1}^{n} (A^{-1})_{ij} b_j.$$

在此等式右端使用(5.29),然后与(5.28)比较,得

$$(A^{-1}b)_i = x_i, \quad i = 1, \cdots, n.$$

此即 $A^{-1}b = x$,这里利用了(5.26),这表明(5.29)定义 A^{-1} 的确是 A 的逆.

现设 A 可逆. A^{-1} 满足

$$A^{-1}A = I,$$

其中 $I = [e_1 \quad \cdots \quad e_n]$ 是单位矩阵.对此等式取行列式,依 **1.5.8** 及 **1.5.2** 的(3),得

$$(\det A^{-1})(\det A) = \det I = 1.$$

因此, $\det A \neq 0$. □

一般,设 $A = [a_{ij}]$ 是 $n \times n$ 矩阵,

$$\left[(-1)^{i+j} \det A_{ij}\right]$$

是以 a_{ij} 的余子式为 (i, j)-元素的矩阵,其转置矩阵

$$\mathrm{adj}A \equiv \left[(-1)^{i+j}\det A_{ji}\right], \tag{5.30}$$

称为 A 的**伴随矩阵**(adjoint matrix).

不巧,"伴随矩阵"这一术语有重叠(见 **1.7.19**).但是只要注意使用场合是不难加以区分的.

现在,(5.29)可以写成

$$A^{-1} = \frac{1}{\det A}\mathrm{adj}A. \tag{5.31}$$

容易证明(留作练习):

(1) $\mathrm{adj}(A)^{\mathrm{T}} = (\mathrm{adj}A)^{\mathrm{T}}$.

(2) $\mathrm{adj}\overline{A} = \overline{\mathrm{adj}A}$.

(3) $\mathrm{adj}I = I$.

(4) $\mathrm{adj}(\alpha A) = \alpha^{n-1}\mathrm{adj}A$.

对于任何置换 p,因符号差成立(证明留作练习),则

$$\sigma(p) = \sigma(p^{-1}). \tag{5.32}$$

并利用公式(5.12),容易证明

$$\det A^{\mathrm{T}} = \det A. \tag{5.33}$$

1.5.14 定义 设 p 是 n 个对象的置换,相伴的一个矩阵 P 如下:

$$(P)_{ij} = \begin{cases} 1, & j = p(i), \\ 0, & j \neq p(i), \end{cases} \quad i, j = 1, \cdots, n. \tag{5.34}$$

P 称为**置换矩阵**(permutation matrix).

容易证明:P 作用于任何向量 x 是对 x 的分量实现置换 p;若 p,q 是两个置换,P,Q 是相伴的置换矩阵,则复合 $p \circ q$ 的相伴矩阵是 PQ.

1.5.15 定义 $n\times n$ 矩阵 $A=[a_{ij}]$ 的**迹**,记作 $\mathrm{tr}A$,是指 A 的对角元素之和:

$$\mathrm{tr}A=\sum_{i=1}^{n}a_{ii}\ . \tag{5.35}$$

行列式是矩阵的一个重要的数量函数.迹是矩阵的另一个重要的数量函数.

1.5.16 定理 迹有下列两条性质:

(1) 线性:对任何 $n\times n$ 矩阵 A 及任何数 α,成立

$$\mathrm{tr}(\alpha A)=\alpha\mathrm{tr}A\ .$$

对任何 $n\times n$ 矩阵 A 和 B,成立

$$\mathrm{tr}(A+B)=\mathrm{tr}A+\mathrm{tr}B\ .$$

(2) 交换性:对任何 $n\times n$ 矩阵 A 和 B,成立

$$\mathrm{tr}(AB)=\mathrm{tr}(BA)\ . \tag{5.36}$$

证 依(5.35),线性是显然的.

根据矩阵乘法,

$$(AB)_{ii}=\sum_{k=1}^{n}a_{ik}b_{ki},\quad (BA)_{ii}=\sum_{k=1}^{n}b_{ik}a_{ki}\ .$$

因此

$$\mathrm{tr}(AB)\equiv\sum_{i=1}^{n}\sum_{k=1}^{n}a_{ik}b_{ki}=\sum_{i=1}^{n}\sum_{k=1}^{n}b_{ik}a_{ki}\equiv\mathrm{tr}(BA)\ ,$$

其中只须交换一次指标 i 和 k 的命名. □

关系式(5.36)可以推广至 A 是 $m\times n$ 矩阵和 B 是 $n\times m$ 矩阵的情形(证明留作练习).

容易证明:若 A 是 $n\times n$ 矩阵, A^{T} 是其转置矩阵,则

$$\mathrm{tr}(AA^{\mathrm{T}})=\sum_{i,j=1}^{n}a_{ij}^{2}\ . \tag{5.37}$$

1.5.17 定义 $n\times n$矩阵A称为**相似**于$n\times n$矩阵B,如果存在一个可逆的$n\times n$矩阵S,使得

$$A=SBS^{-1}. \tag{5.38}$$

设S是可逆的$n\times n$矩阵.对于任意的$n\times n$矩阵B,映射

$$B\to SBS^{-1} \tag{5.39}$$

称为以S为**相似矩阵**(similar matrix)的**相似变换**.

这一定义平行于**1.3.11**中关于线性映射相似的定义(3.36).

相似性是一种等价关系,即具有

(1) 自反性: A与自己相似.

(2) 对称性: 若A相似于B,则B相似于A.

(3) 传递性: 若A相似于B,B相似于C,则A相似于C.

证明如下:

对于(1),只要在(5.38)中取$A=B,S=I$.

对于(2),对(5.38)右乘S,左乘S^{-1},得

$$B=S^{-1}AS=(S^{-1})A(S^{-1})^{-1}.$$

对于(3),设

$$A=SBS^{-1},B=TCT^{-1},$$

则利用结合律得

$$A=SBS^{-1}=S(TCT^{-1})S^{-1}=(ST)C(ST)^{-1}.$$

这就证明了A相似于C.

1.5.18 定理 相似的矩阵的行列式相等,而且有相同的迹.

证 依**1.5.8**,从(5.38)得

$$\begin{aligned}\det A&=(\det S)(\det B)(\det S^{-1})\\&=(\det B)(\det S)(\det S^{-1})\\&=(\det B)\det(SS^{-1})=(\det B)\det I\\&=\det B.\end{aligned}$$

依 1.5.16 的(2),则

$$\begin{aligned}\mathrm{tr}A &= \mathrm{tr}(SBS^{-1}) = \mathrm{tr}((SB)S^{-1}) \\ &= \mathrm{tr}(S^{-1}(SB)) = \mathrm{tr}B.\end{aligned}$$ □

在 **1.4.5** 中论述过:从一个 n 维线性空间 X 到其自身的任何线性映射 T,通过在 X 中选取基,可以表示作 $n\times n$ 矩阵;产生自两种基的不同选取的两种不同表示的矩阵是相似的.因此,鉴于上述定理,可以定义如此线性映射 T 的行列式和迹为表示 T 的矩阵的行列式和迹.

1.6 谱论

1.6.1 引言 谱论是分析从空间到其自身的线性映射,目的是将此类映射分解成它们的基本组成部分.

先看一个源自周期运动稳定性的问题.

假定所研究的系统的状态可以用 n 个参数来描述,这些参数归并成 $\mathbb{R}^n$ 中单个向量 u;而且只要系统的初始状态已知,系统关于时间 t 的控制发展定律唯一确定系统在后面任何时间的状态.

用 $u(0)$ 表示系统在 $t=0$ 时的状态;因而 $t=1$ 时其状态完全由 $u(0)$ 确定,记之为 $u(1)\equiv F(u(0))$.假定 F 是一个可微函数,且系统的控制发展定律无论何时是同一的,因此系统在 $t=2$ 时状态为 $u(2)\equiv F(u(1))$.更一般地,F 是由系统在时间 t 的状态到在时间 $t+1$ 的状态的对应关系.

假定运动具有周期性,周期为 1,并且在时间 $t=0$ 时的初始状态 $u(0)=0$.于是,当 $t=1$ 时状态回到 0,即

$$F(0)=0, \tag{6.1}$$

这种运动称为**稳定**的,如果从充分接近于零(原点)的任何点 x 出发,

运动随 t 趋于无穷而趋于零.

因为描述运动的函数 F 是可微的,所以对于小的 x, $F(x)$ 可用线性逼近来刻画:

$$F(x) \approx Ax.$$

针对这里要讨论的目的,直接假定 F 是线性函数

$$F(x) = Ax, \tag{6.2}$$

其中 A 是 $n \times n$ 矩阵.从点(状态) x 出发,经过 m 个单位时间后,系统处于位置(状态)

$$A^m x. \tag{6.3}$$

下面考察如下形式的序列:

$$x, Ax, \cdots, A^m x, \cdots. \tag{6.4}$$

先看几个矩阵之幂 $A^m x$ 的性态的例子;取 $m = 1024$,这样,可以通过执行10次平方运算求得 A^m:

	(a)	(b)	(c)	(d)
A:	$\begin{bmatrix} 3 & 2 \\ 1 & 4 \end{bmatrix}$	$\begin{bmatrix} 3 & 2 \\ -5 & -3 \end{bmatrix}$	$\begin{bmatrix} 5 & 7 \\ -3 & -4 \end{bmatrix}$	$\begin{bmatrix} 5 & 6.9 \\ -3 & -4 \end{bmatrix}$
A^{1024}:	$> 10^{700}$	$\begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$	$\begin{bmatrix} -5 & -7 \\ 3 & 4 \end{bmatrix}$	$< 10^{-78}$

这些数值实验有力地提示:

(1) $A^m \to \infty$,当 $m \to \infty$.

(2) $A^m = I$,当 $m = 1024$.

(3) $A^m = -A$,当 $m = 1024$.

(4) $A^m \to 0$,当 $m \to \infty$;即 A^m 的每个元素趋于零.

现在转向形如(6.4)的序列的性态分析.假定向量 $x \neq 0$ 关于矩阵 A 存在数 λ,成立

$$Ax = \lambda x, \tag{6.5}$$

则

$$A^m x = \lambda^m x. \tag{6.6}$$

在如此情形下，序列(6.4)的性态如下:

(1) 若 $|\lambda|>1$,则 $A^m x \to \infty$,当 $m \to \infty$.

(2) 若 $|\lambda|<1$,则 $A^m x \to 0$,当 $m \to \infty$.

(3) 若 $\lambda = 1$,则 $A^m x = x$,对所有 m.

1.6.2 定义 设 A 是 $n \times n$ 矩阵,向量 $x \neq 0$ 和数 λ 满足(6.5).则称 λ 为 A 的**特征值**,称 x 为 A 的相应特征值 λ 的**特征向量**.

对于 $n \times n$ 矩阵来说,特征值和特征向量是不是一种普遍的概念呢?为了搞清这一问题,把(6.5)改写成如下形式:

$$(\lambda I - A)x = 0, \tag{6.7}$$

这说明 $x \neq 0$ 属于 $\lambda I - A$ 的零空间,因此矩阵 $\lambda I - A$ 不可逆.这样,依 **1.5.13**, A 存在特征向量的充分必要条件是

$$\det(\lambda I - A) = 0. \tag{6.8}$$

同样,(6.8)也是 λ 为 A 的特征值的充分必要条件.只须证明充分性.

事实上,若(6.8)成立,则矩阵 $\lambda I - A$ 不可逆;依 **1.3.4**,存在 $x \neq 0$ 属于 $\lambda I - A$ 的零空间,即存在 $x \neq 0$ 成立(6.7),从而 x 是特征向量, λ 是特征值.

1.6.3 定义 设 A 是 $n \times n$ 矩阵.方程(6.8)称为 A 的**特征方程**;其左边是关于 λ 的多项式,称为 A 的**特征多项式**,记作 $p_A(\lambda)$,则

$$p_A(\lambda) \equiv \det(\lambda I - A). \tag{6.9}$$

p_A 是 n 次多项式,其最高幂 λ^n 的系数为 1.根据代数学基本定理, n 次复系数多项式有 n 个复根,有的根可以是多重的.特征多

项式的根是 A 的特征值.为了保证如此多项式有一组完整的根,线性映射的谱论是在复数域的线性空间中阐述的;因而,下面将主要针对复矩阵 $A\in\mathbb{C}^{n\times n}$ 来讨论.

1.6.4 定理 矩阵 $A\in\mathbb{C}^{n\times n}$ 的相应于不同特征值的特征向量是线性无关的.

证 设

$$\lambda_i\neq\lambda_k,\quad i\neq k,\quad i,k=1,\cdots,m.$$

而且

$$Ax^{(i)}=\lambda_i x^{(i)},\quad x^{(i)}\neq 0,\quad i=1,\cdots,m. \tag{6.10}$$

这里 $m\leq n$.现在假设特征向量 $x^{(1)},\cdots,x^{(m)}$ 之中存在非平凡线性关系.可以有若干种非平凡线性关系;因 $x^{(1)},\cdots,x^{(m)}$ 全不为零,故每种关系至少包含两个特征向量.于是,在这些关系中必有包含特征向量个数最少的,设其包含的特征向量个数为 $s,s\leq m$,不妨设 $x^{(1)},\cdots,x^{(s)}$ 存在如下关系:

$$\sum_{j=1}^{s}\alpha_j x^{(j)}=0;\quad \alpha_j\neq 0,j=1,\cdots,s. \tag{6.11}$$

将 A 作用于(6.11)并利用(6.10),得

$$\sum_{j=1}^{s}\alpha_j Ax^{(j)}=\sum_{j=1}^{s}\alpha_j\lambda_j x^{(j)}=0, \tag{6.12}$$

用 λ_s 乘以(6.11)而后与(6.12)相减,有

$$\sum_{j=1}^{s}\alpha_j(\lambda_j-\lambda_s)x^{(j)}=\sum_{j=1}^{s-1}\alpha_j(\lambda_j-\lambda_s)x^{(j)}=0.$$

显然,这是仅包含 $s-1$ 个特征向量 $x^{(1)},\cdots,x^{(s-1)}$ 的非平凡线性关系,与 s 是满足如此关系的最小的特征向量个数矛盾. □

1.6.5 定理 若矩阵 $A\in\mathbb{C}^{n\times n}$ 的特征多项式有 n 个不同的根,则 A 有 n 个线性无关的特征向量.

此定理是 **1.6.4** 的直接推论.

在矩阵 $A \in \mathbb{C}^{n\times n}$ 有 n 个不同特征值 $\lambda_1, \cdots, \lambda_n$ 的情形,其任何 n 个线性无关的特征向量 $x^{(1)}, \cdots, x^{(n)}$ 构成 $\mathbb{C}^n$ 的一组基;因此 $\mathbb{C}^n$ 的每一向量 x 可以表示为 $x^{(1)}, \cdots, x^{(n)}$ 的线性组合:

$$x = \sum_{j=1}^{n} \alpha_j x^{(j)}, \tag{6.13}$$

将 A^m 作用于(6.13),得

$$A^m x = \sum_{j=1}^{n} \alpha_j \lambda_j^m x^{(j)}. \tag{6.14}$$

由此容易证明:

(1) 当所有特征值按模小于1时,对 $\mathbb{C}^n$ 中所有向量 x,有

$$A^m x \to 0, \quad 当 m \to \infty,$$

即 $A^m x$ 的所有分量趋向零.

(2) 当所有特征值按模大于1时,对 $\mathbb{C}^n$ 中所有向量 $x \neq 0$,有

$$A^m x \to \infty, \quad 当 m \to \infty,$$

即 $A^m x$ 的某些分量趋向无穷.

1.6.6 定理 设 $\lambda_1, \cdots, \lambda_n$ 是矩阵 $A \in \mathbb{C}^{n\times n}$ 的特征值,它们作为 A 的特征方程(6.8)的根,重根按重数计.则

$$\sum_{i=1}^{n} \lambda_i = \mathrm{tr} A, \quad \prod_{i=1}^{n} \lambda_i = \det A. \tag{6.15}$$

证 根据代数基本定理, A 的特征多项式 p_A 可作因式分解

$$\begin{aligned} p_A(\lambda) &= \prod_{i=1}^{n} (\lambda - \lambda_i) \\ &= \lambda^n - \left(\sum_{i=1}^{n} \lambda_i \right) \lambda^{n-1} + \cdots + (-1)^n \prod_{i=1}^{n} \lambda_i . \end{aligned} \tag{6.16}$$

另一方面,利用(5.12)将行列式表示成乘积之和:

$$
\begin{aligned}
p_A(\lambda) &= \det(\lambda I - A) \\
&= \det\begin{bmatrix} \lambda - a_{11} & -a_{12} & \cdots & -a_{1n} \\ -a_{21} & \lambda - a_{22} & \cdots & -a_{2n} \\ \cdots & \cdots & \cdots & \cdots \\ -a_{n1} & -a_{n2} & \cdots & \lambda - a_{nn} \end{bmatrix} \\
&= \sum_p \sigma(p)\left[\prod_{i=1}^{n}(\lambda\delta_{p,i} - a_{p,i})\right],
\end{aligned}
$$

这里 p 是 $1,\cdots,n$ 这 n 个数的任意的置换; δ_{ij} 是 Kronecker 符号:

$$
\delta_{ij} = \begin{cases} 1, & i = j, \\ 0, & i \neq j, \end{cases} \quad i, j = 1,\cdots,n.,
$$

显然, λ 的 n 次项和 $n-1$ 次项是单一的从对角元素的乘积

$$
\prod_{i=1}^{n}(\lambda - a_{ii}) = \lambda^n - (\operatorname{tr} A)\lambda^{n-1} + \cdots
$$

产生的,即此式右端前两项等同于(6.16)相应的项; p_A 的常数项等于

$$
p_A(0) = \det(-A) = (-1)^n \det A,
$$

从(6.16)看,又有

$$
p_A(0) = (-1)^n \prod_{i=1}^{n} \lambda_i . \qquad \square
$$

1.6.7 谱映射定理 设 $A \in \mathbb{C}^{n\times n}$, q 是任意的多项式.则

(1) 如果 λ 是 A 的特征值,那么 $q(\lambda)$ 是 $q(A)$ 的特征值.

(2) $q(A)$ 的每个特征值形如 $q(\lambda)$,其中 λ 是 A 的特征值.

证 设 $q(t) = \sum_{i=0}^{m} c_i t^i$.

(1) 用 x 表示 A 的相应于特征值 λ 的特征向量.利用(6.6),

$$q(A)x=\sum_{i=0}^{m}c_iA^ix=\sum_{i=0}^{m}c_i\lambda^ix=q(\lambda)x. \tag{6.17}$$

这说明 $q(\lambda)$ 是 $q(A)$ 的特征值,而且 $q(A)$ 与 A 以 x 为共同的特征向量.

(2) 用 μ 表示 $q(A)$ 的一个特征值,这意味着 $q(A)-\mu I$ 不可逆.设多项式 $q(t)-\mu$ 的因式分解为

$$q(t)-\mu=c_m\prod_{i=1}^{m}(t-\nu_i),$$

于是

$$q(A)-\mu I=c_m\prod_{i=1}^{m}(A-\nu_iI).$$

在此等式中,因为左端 $q(A)-\mu I$ 不可逆,所以右端至少有一个因子 $A-\nu_rI$ 不可逆.这就是说,存在 $\lambda\equiv\nu_r$ 是 A 的特征值.注意到此 λ 是 $q(t)-\mu$ 的一个根,知 $\mu=q(\lambda)$. □

1.6.8 引理 设 P 和 Q 是两个以矩阵为系数的多项式, 即

$$P(t)\equiv\sum_j P_jt^j,\quad Q(t)\equiv\sum_k Q_kt^k,$$

积 $R\equiv PQ$,则

$$R(t)=\sum_l R_lt^l,\quad R_l=\sum_{j+k=l}P_jQ_k.$$

而且,当矩阵 A 与 Q 的每个系数矩阵 Q_k 可交换的情况下,成立

$$P(A)Q(A)=R(A). \tag{6.18}$$

此引理是不证自明的.

1.6.9 Cayley-Hamilton 定理 每个矩阵 $A\in\mathbb{C}^{n\times n}$ 满足其自己的特征方程:

$$p_A(A)=0. \tag{6.19}$$

证 若 A 有 n 个不同的特征值 $\lambda_1,\cdots,\lambda_n$，则有 n 个线性无关的特征向量 $x^{(1)},\cdots,x^{(n)}$. 将 $p_A(A)$ 作用于(6.13)，得到对 $\mathbb{C}^n$ 中每一向量 x 成立

$$p_A(A)x=\sum_{j=1}^{n}\alpha_j p_A(\lambda_j)x^{(j)}=\sum_{j=1}^{n}0=0\,. \qquad (6.20)$$

由此推出，对有 n 个不同特征值的情形成立(6.19).

下面进一步证明(6.19)对所有矩阵成立. 设

$$Q(\lambda)\equiv \lambda I-A\,,$$

而 $\mathrm{adj}Q(\lambda)$ 是(5.30)定义的 $Q(\lambda)$ 的伴随矩阵. 根据(5.31),

$$p_A(\lambda)\equiv(\det Q(\lambda))I=(\mathrm{adj}Q(\lambda))Q(\lambda)\,. \qquad (6.21)$$

因为 A 与 $Q(\lambda)$ 的系数矩阵可交换，所以依 **1.6.8**，在(6.21)中可以置 $\lambda=A$；然后注意到 $Q(A)=0$，即得(6.19). □

1.6.10 定义 设对于矩阵 $A\in\mathbb{C}^{n\times n}$ 的特征值 λ，存在向量 x，满足 $x\neq 0$，且对某正整数 m 成立

$$(A-\lambda I)^m x=0\,, \qquad (6.22)$$

则称 x 为 A 的属于特征值 λ 的**广义特征向量**.

对于矩阵 $A\in\mathbb{C}^{n\times n}$，从 **1.6.5** 知道，当 A 的特征方程没有重根时，存在 n 个线性无关的特征向量；然而，当 A 的特征方程有重根时，不能一般地期望也存在 n 个线性无关的特征向量.

例 (1) $A\equiv I\in\mathbb{C}^{n\times n}$ 是单位矩阵. 特征方程

$$p_A(\lambda)\equiv(\lambda-1)^n=0\,,$$

1 是其 n 重根. 此情形，每个非零向量 $x\in\mathbb{C}^n$ 都是 A 的特征向量；这蕴涵 A 存在 n 个线性无关的特征向量.

(2) $A=\begin{bmatrix}3 & 2\\ -2 & -1\end{bmatrix}$. 特征方程

$$p_A(\lambda)\equiv(\lambda-1)^2=0,$$

1 是其 2 重根.此情形,A 的特征向量 $x\equiv(x_1,x_2)^{\mathrm{T}}$ 应满足方程 $Ax=x$,即

$$\begin{pmatrix}3x_1+2x_2\\-2x_1-x_2\end{pmatrix}=\begin{pmatrix}x_1\\x_2\end{pmatrix}.$$

此方程等价于方程 $x_1+x_2=0$;从而,A 的所有特征向量形如

$$c\begin{pmatrix}1\\-1\end{pmatrix},$$

其中 c 是任意的非零常数,这表明 A 不存在两个线性无关的特征向量.

广义特征向量概念是特征向量概念的推广,以弥补 $n\times n$ 矩阵不是普遍存在 n 个线性无关的特征向量的缺陷.为了推证出这样的结论,需要下面两个代数学的结果.

1.6.11 引理 设 p 和 q 是一对复系数多项式而且没有共同的零点,则存在两个多项式 a 和 b,使得

$$ap+bq\equiv1. \tag{6.23}$$

证 用 $\mathscr{P}$ 表示所有形如 $ap+bq$ 的多项式,其中必存在次数最低的非零多项式,取其一记之为 d.

现在,首先证明 p 和 q 均可被 d 整除.如若不然,则除法算式产生一个余项 $r\neq0$,

$$r=p-md.$$

因 p 和 d 属于 $\mathscr{P}$,故 $p-md=r$ 也属于 $\mathscr{P}$.然而 r 的次数低于 d 的次数,这是一个矛盾.

其次,证明 d 的次数为零.事实上,如果 d 的次数大于零,那么根据代数基本定理,d 至少有一个根.因 d 整除 p 和 q,故 p 和 q 有共同的根,但这与假设是矛盾的.

因此,$d\equiv\text{const.}\neq0$,特别可取 $d\equiv1$. □

1.6.12 引理 设 p 和 q 是一对复系数多项式且没有共同的零点, $A\in\mathbb{C}^{n\times n}$, $\mathscr{N}_p,\mathscr{N}_q,\mathscr{N}_{pq}$ 分别是 $p(A),q(A),p(A)q(A)$ 的零空间, 则 $\mathscr{N}_{pq}$ 是 $\mathscr{N}_p$ 与 $\mathscr{N}_q$ 的直和:

$$\mathscr{N}_{pq}=\mathscr{N}_p\oplus\mathscr{N}_q, \tag{6.24}$$

或者等价地说,每个 $x\in\mathscr{N}_{pq}$ 可以唯一地分解成

$$x=x_p+x_q,\quad x_p\in\mathscr{N}_p,\quad x_q\in\mathscr{N}_q, \tag{6.25}$$

一般,设 $p_1,\cdots,p_k$ 是两两没有共同的零点的多项式,则成立

$$\mathscr{N}_{p_1\cdots p_k}=\mathscr{N}_{p_1}\oplus\cdots\oplus\mathscr{N}_{p_k}. \tag{6.26}$$

证 利用(6.23),得

$$a(A)p(A)+b(A)q(A)=I. \tag{6.27}$$

因此

$$a(A)p(A)x+b(A)q(A)x=x. \tag{6.28}$$

现设 $x\in\mathscr{N}_{pq}$.由于同一矩阵的多项式之间的可交换性,以及 x 属于 $p(A)q(A)$ 的零空间,

$$q(A)a(A)p(A)x=a(A)p(A)q(A)x=0.$$

这样,证明了(6.28)左端第一项属于 $q(A)$ 的零空间 $\mathscr{N}_q$,类似地可证(6.28)左端第二项属于 $p(A)$ 的零空间 $\mathscr{N}_p$.因此,(6.28)就是形如(6.25)的分解式.

为了证明分解的唯一性,假设

$$x=x_p+x_q=x'_p+x'_q,$$

则

$$y\equiv x_p-x'_p=x'_q-x_q$$

是同时属于 $\mathscr{N}_p$ 和 $\mathscr{N}_q$ 的一个元素,利用(6.27),

$$a(A)p(A)y+b(A)q(A)y=y,$$

此式左边两项同时为零,故 $y=0$.所以

$$x_p=x'_p,\quad x_q=x'_q.$$ □

1.6.13 谱定理 设$A\in\mathbb{C}^{n\times n}$,则$\mathbb{C}^n$中的每个向量可以写成$A$的真的或广义的特征向量之和.

证 设$x\in\mathbb{C}^n$是任一向量,因为$n+1$个向量

$$x, Ax, A^2x, \cdots, A^nx$$

必线性相关,所以存在次数小于等于n的多项式p,并将其因式分解,使得

$$p(A)x\equiv\prod_{j=1}^{k}(A-r_jI)^{m_j}x=0, \tag{6.29}$$

其中r_j是p的根,m_j是其重数.当r_j不是A的特征值时,$A-r_jI$可逆.在(6.29)中,因各因式可交换,故可将所有可逆因式去除,从而剩下r_j全是A的特征值.不妨假设p就是已去除可逆因式的多项式,记

$$p_j(t)=(t-r_j)^{m_j}, \quad j=1,\cdots k. \tag{6.30}$$

则(6.29)可以写成

$$\prod p_j(A)x=0.$$

也就是说,$x\in\mathcal{N}_{p_1\cdots p_k}$.显然, $p_1,\cdots,p_k$两两没有共同零点,因此能够应用 **1.6.12**, x必可分解成$\mathcal{N}_{p_1},\cdots,\mathcal{N}_{p_k}$中的向量$x_1,\cdots,x_k$之和;然而,根据(6.30)并依 **1.6.10**, $\mathcal{N}_{p_j}$中的每个向量是A的广义特征向量. □

1.6.14 定义 设$A\in\mathbb{C}^{n\times n}$,用$\mathscr{P}$表示满足$p(A)=0$的多项式$p$的全体,$\mathscr{P}$中次数最低且首项系数为1的非零多项式称为$A$的**最小多项式**,记作$m_A$.

显然,$\mathscr{P}$中任何两个多项式之和仍属于$\mathscr{P}$;而且,若p属于$\mathscr{P}$,则p乘任一多项式也属于$\mathscr{P}$.

容易证明: A 的最小多项式 m_A 是唯一的.事实上,m_A 可以整除 $\mathscr{P}$ 中的所有 p.如若不然,则除法过程 $p = qm_A + r$ 给出比 m_A 次数低的非零多项式 r.显然,$r = p - qm_A$ 属于 $\mathscr{P}$,但是这与 m_A 是 $\mathscr{P}$ 中次数最低的非零多项式矛盾.这样,$\mathscr{P}$ 中次数最低的非零多项式与 m_A 至多差一常数因子,而若进一步固定首项系数为1便唯有 m_A 了.

1.6.15 定义 设 $A \in \mathbb{C}^{n\times n}$,$\lambda$ 是 A 的特征值.由广义特征向量组成的 $(A-\lambda I)^m$ 的零空间记作 $\mathscr{N}_m \equiv \mathscr{N}_m(\lambda)$.显然,$\mathscr{N}_m$ 随指数 m 增大有可能扩充而不会缩减,而且因是有限维空间的子空间,必存在指数 $d \equiv d(\lambda)$,使得

$$\mathscr{N}_1 \subset \cdots \subset \mathscr{N}_{d-1} \subset \mathscr{N}_d = \mathscr{N}_{d+1} = \cdots, \quad \mathscr{N}_{d-1} \neq \mathscr{N}_d . \tag{6.31}$$

如此 $d(\lambda)$ 称为 A 的特征值 λ 的**指数**.

1.6.16 定理 设 $A \in \mathbb{C}^{n\times n}$,$\lambda_1, \cdots, \lambda_k$ 是 A 的所有不同的特征值,它们的指数分别为 $d_1, \cdots, d_k$.则

$$m_A(\lambda) = \prod_{i=1}^{k} (\lambda - \lambda_i)^{d_i}, \tag{6.32}$$

m_A 是 A 的最小多项式.

证 根据假设,以及 **1.6.13** 的 $\mathbb{C}^n$ 中每个向量可以写成 A 的真的或广义的特征向量之和的结论,即知对(6.32)给定的 m_A 成立

$$m_A(A)x = \prod_{i=1}^{k} (A - \lambda_i I)^{d_i} x = 0, \quad \forall x \in \mathbb{C}^n .$$

由此,m_A 满足 $m_A(A) = 0$.

另一方面,设 A 的最小多项式为 p,从 **1.6.14** 中的讨论知 p 必整除(6.32)给定的 m_A,于是 p 形如

$$p(\lambda) = \prod_{i=1}^{k} (\lambda - \lambda_i)^{d_i'}; \quad d_i' \le d_i, \quad i = 1, \cdots, k .$$

如果 p 的次数低于 m_A 的次数,那么存在某 $d'_j < d_j$.依特征值的指数定义 **1.6.15**,有

$$\mathscr{N}_{d'_j}(\lambda_j) \subset \mathscr{N}_{d_j}(\lambda_j), \quad \mathscr{N}_{d'_j}(\lambda_j) \neq \mathscr{N}_{d_j}(\lambda_j).$$

这意味着存在 $y \in \mathbb{C}^n$:

$$y \in \mathscr{N}_{d_j}(\lambda_j) \quad 但 \quad y \notin \mathscr{N}_{d'_j}(\lambda_j).$$

或者说

$$(A - \lambda_j I)^{d'_j} y \neq 0 \quad 但 \quad (A - \lambda_j I)^{d_j} y = 0.$$

从而 $p(A)y \neq 0$ 并推出 $p(A) \neq 0$,这与 p 是 A 的最小多项式矛盾.因此 p 与 m_A 只能差一常数因子.

另外,(6.32)给定的 m_A 的首项系数显然为1. □

现在,在定理 **1.6.16** 的条件下,谱定理 **1.6.13** 可表述为

$$\mathbb{C}^n = \mathscr{N}^{(1)} \oplus \cdots \oplus \mathscr{N}^{(k)}, \tag{6.33}$$

这里

$$\mathscr{N}^{(i)} \equiv \mathscr{N}_{d_i}(\lambda_i), \quad i = 1, \cdots, k.$$

称它们为 A 的**广义特征空间**.

$\mathscr{N}^{(i)}$ 的维数等于 λ_i 作为 A 的特征方程的根的重数.这一命题的证明用到微分学,见 **9.5.18**.

容易证明, A 将每个子空间 $\mathscr{N}^{(i)}$ 映入 $\mathscr{N}^{(i)}$,这样的子空间称为在 A 下是**不变**的.

1.6.17 引理 设 $A \in \mathbb{C}^{n\times n}$, $\lambda = 0$ 是 A 的特征值,并且用(6.31)表示与其相应的零空间的序列, d 是其指数,则 A 将 $(\mathscr{N}_{m+1} / \mathscr{N}_m)$ 映入 $(\mathscr{N}_m / \mathscr{N}_{m-1})$,而且是一对一的.

证 所要证明的结论几乎是"$\mathscr{N}_{j+1}$ 是 $\mathscr{N}_j$ 在 A 下的逆象"这一事实的直接推论. □

1.6.18 定理　设 $A,B\in\mathbb{C}^{n\times n}$.

(1) 若 A 和 B 相似,

$$A=SBS^{-1}, \tag{6.34}$$

S 是某可逆矩阵,则 A 和 B 有相同的特征值 $\lambda_1,\cdots,\lambda_k$,而且成立

$$\dim\mathscr{N}_m(\lambda_j)=\dim\mathscr{M}_m(\lambda_j),$$
$$j=1,\cdots,k,\quad m=1,2,\cdots, \tag{6.35}$$

这里 $\mathscr{N}_m(\lambda_j)$ 是 $(A-\lambda_jI)^m$ 的零空间,$\mathscr{M}_m(\lambda_j)$ 是 $(B-\lambda_jI)^m$ 的零空间.

(2) 反之,若 A 和 B 有相同的特征值 $\lambda_1,\cdots,\lambda_k$,而且成立条件(6.35),则 A 和 B 相似.

证　(1)是明显的:因为若 A 和 B 相似,则 $A-\lambda I$ 和 $B-\lambda I$ 也相似,从而它们的任何同次幂均相似:

$$(A-\lambda I)^m=S(B-\lambda I)^mS^{-1}, \tag{6.36}$$

而两个相似矩阵的零空间有相同的维数.由此推出 A 和 B 有相同的特征值,而且成立(6.35).

为了证明逆命题(2),从关系式(6.35)出发构造确立 A 和 B 相似的映射 S.关系式(6.36)提示怎样去构造 S,即必须成立

$$\mathscr{N}_m(\lambda_j)=S\mathscr{M}_m(\lambda_j),\quad j=1,\cdots,k,\quad m=1,2,\cdots. \tag{6.37}$$

下面固定 j 并将其省略,而且假定 $\lambda=\lambda_j=0$,这可以通过事先从 A 和 B 两者减去 λ_jI 来实现.设 d 是 $\lambda=0$ 的指数,记

$$\mathscr{N}_m\equiv\mathscr{N}_m(0),\quad \mathscr{M}_m\equiv\mathscr{M}_m(0).$$

考虑零空间序列

$$\mathscr{N}_1\subset\cdots\subset\mathscr{N}_d,$$

在 $\mathscr{N}_d$ 中引出一组特殊的基,其元素按束组成.第一束

$$x^{(1)},\cdots,x^{(l)},$$

其中 $l\equiv\dim(\mathscr{N}_d/\mathscr{N}_{d-1})$,它们是任何 l 个 $\operatorname{mod}\mathscr{N}_{d-1}$ 线性无关的向量.下一束形如

$$Ax^{(1)},\cdots,Ax^{(l)},$$

依 **1.6.17**,它们在 $\mathscr{N}_{d-1}$ 之中;它们是 $\bmod \mathscr{N}_{d-2}$ 线性无关的,是 $\mathscr{N}_{d-1}/\mathscr{N}_{d-2}$ 的一组基.现在重复这样的过程,再对 $Ax^{(1)},\cdots,Ax^{(l)}$ 作用以 A,得到的一束向量是 $\mathscr{N}_{d-2}/\mathscr{N}_{d-3}$ 的一组基;如此,一直继续至到达 $\mathscr{N}_1$.

按照逆命题(2)的假设,矩阵 B 相应的零空间序列

$$\mathscr{M}_1 \subset \cdots \subset \mathscr{M}_d$$

中每个 $\mathscr{M}_m$ 和相应的 $\mathscr{N}_m$ 有相同的维数.所以上述构造特殊基的步骤适合 $\mathscr{M}_d$ 引出一组基,而且其每束和 $\mathscr{N}_d$ 相应的束有同样多个基元素.指定 $\mathscr{M}_d$ 中的每个基元素 $y^{(i)}$ 对应于相应的束中的基元素 $x^{(i)}$,而且使得 $Ax^{(i)}$ 对应于 $By^{(i)}$,因为维数相匹配,这一点是可以做到的. $\mathscr{M}_d$ 的基元素到 $\mathscr{N}_d$ 的基元素的一对一的指派一经确立,便可唯一地将这个指派延拓成 $\mathscr{M}_d$ 到 $\mathscr{N}_d$ 上的一个线性映射 S.显然,在 $\mathscr{M}_d$ 上成立

$$AS = SB. \tag{6.38}$$

根据(6.33),$\mathbb{C}^n$ 是 $\mathscr{M}_{d_1}^{(1)},\cdots,\mathscr{M}_{d_k}^{(k)}$ 的直和,也是 $\mathscr{N}_{d_1}^{(1)},\cdots,\mathscr{N}_{d_k}^{(k)}$ 的直和.因此,S 可以延拓成 $\mathbb{C}^n$ 到 $\mathbb{C}^n$ 上的可逆映射,使得(6.38)成立.这就证明了 A 和 B 相似. □

1.6.13, 1.6.16 和 **1.6.18** 是矩阵谱论的基本事实.值得指出的是,进入这些定理的概念——特征值、特征向量、广义特征向量、指数——对于任何复数域$\mathbb{C}$上有限维线性空间 X 到其自身的任一映射仍然是有意义的.

1.6.19 定理 设 X 是复数域$\mathbb{C}$上有限维线性空间,A 和 B 是 X 到其自身的线性映射,且可交换:

$$AB = BA, \tag{6.39}$$

则 X 中存在同为 A 和 B 的特征向量和广义特征向量组成的基.

证 依 **1.6.13** 和关系式(6.33), X 可以分解为 A 的广义特征空间的直和:

$$X = \mathscr{N}^{(1)} \oplus \cdots \oplus \mathscr{N}^{(k)},$$

其中 $\mathscr{N}^{(j)}$ 是 $(A-\lambda_j I)^{d_j}$ 的零空间.由于假定 B 可以和 A 交换,有

$$B(A-\lambda I)^d x = (A-\lambda I)^d Bx. \tag{6.40}$$

如果 $x \in \mathscr{N}^{(j)}$ 且 $\lambda = \lambda_j$,那么(6.40)左端为零,从而其右端也为零,表明 $Bx \in \mathscr{N}^{(j)}$.因此, B 映射 $\mathscr{N}^{(j)}$ 到 $\mathscr{N}^{(j)}$.

现在对作为在 $\mathscr{N}^{(j)}$ 上线性映射的 B 应用 **1.6.13** 和关系式(6.33),进一步进行每个 $\mathscr{N}^{(j)}$ 关于 B 的谱分解,便达到所要证明的结论. □

推论 定理 **1.6.19** 可以推广到任意多个两两可交换的映射. □

1.6.20 定理 每个矩阵 $A \in \mathbb{C}^{n\times n}$ 相似于它的转置 A^{T}.

证 因为 A 和 A^{T} 两者的行列式相等,而 $I^{\mathrm{T}} = I$,所以

$$\det(\lambda I - A) = \det(\lambda I - A^{\mathrm{T}}).$$

这就是说, A 和 A^{T} 有相同的特征多项式,从而两者有相同的特征值.

当 A 的特征值各不相同时,依 **1.6.5**, A 有 n 个线性无关的特征向量,从而每个特征值的指数均为1,相应的每个零空间是1维的,此时 A^{T} 也如此.于是从 **1.6.18** 推出 A 和 A^{T} 相似.

当 A 有重特征值和广义特征向量时,依 **1.3.10**, A 和 A^{T} 对应的零空间有相同的维数,然后再利用 **1.6.18**. □

1.6.21 定理 设 X 是复数域 $\mathbb{C}$ 上有限维线性空间, A 是 X 到 X 的线性映射, X' 是 X 的对偶空间,映射 $A': X' \to X'$ 是 A 的转置.又设 λ 和 μ 是 A 的两个不同的特征值,而且 x 是 A 的相应特征值 λ 的特征向量, l 是 A' 的相应特征值 μ 的特征向量.则 l 和 x 相互零化:

$$(l,x)=0.$$

证 根据转置定义 1.3.8,

$$(A'l,x)=(l,Ax),\quad \forall l\in X',x\in X.$$

如果特别取 x 是 A 的特征向量,l 是 A' 的特征向量:

$$Ax=\lambda x,\quad A'l=\mu l,$$

那么

$$\mu(l,x)=\lambda(l,x).$$

由此,因 $\lambda\neq\mu$,故 (l,x) 必须为零. □

1.6.22 定理 设 $A\in\mathbb{C}^{n\times n}$ 有 n 个不同的特征值 $\lambda_1,\cdots,\lambda_n$. A 的相应的特征向量表示为 $x^{(1)},\cdots,x^{(n)}$, A^{T} 的相应的特征向量表示为 $l^{(1)},\cdots,l^{(n)}$.则

(1) $(l^{(i)},x^{(i)})\neq 0,\quad i=1,\cdots,n$.

(2) 如果 x 关于特征向量有展开式

$$x=\sum_{i=1}^{n}\alpha_i x^{(i)},\tag{6.41}$$

那么

$$\alpha_i=\frac{(l^{(i)},x)}{(l^{(i)},x^{(i)})},\quad i=1,\cdots,n.$$

证 注意到 $x^{(1)},\cdots,x^{(n)}$ 构成 $\mathbb{C}^n$ 的一组基,$l^{(1)},\cdots,l^{(n)}$ 也构成 $\mathbb{C}^n$ 的一组基,将 $x^{(i)}$ 表示成 $l^{(1)},\cdots,l^{(n)}$ 的线性组合:

$$x^{(i)}=\sum_{j=1}^{n}c_j l^{(j)}.$$

利用 1.6.21,

$$(x^{(i)},x^{(k)})=\sum_{j=1}^{n}c_j(l^{(j)},x^{(k)})=c_k(l^{(k)},x^{(k)}).$$

特别,$k=i$ 时,

$$c_i(l^{(i)},x^{(i)})=(x^{(i)},x^{(i)})\neq 0.$$

由此即得(1)的结论.

类似地,将$l^{(k)}$作用于(6.41),即可推出(2)的结论. □

1.6.23 Gerschgorin 定理 设$A=[a_{ij}]\in\mathbb{C}^{n\times n}$.定义圆盘

$$G_i(A)\equiv\left\{z:|z-a_{ii}|\le\sum_{j=1,j\ne i}^{n}|a_{ij}|\right\},\quad i=1,\cdots,n;$$

$$z\in\mathbb{C}. \qquad (6.42)$$

记n个圆盘之并为

$$G(A)\equiv\bigcup_{i=1}^{n}G_i(A). \qquad (6.43)$$

则

(1) A的所有特征值包含在$G(A)$之中.或者说,A的每个特征值必包含在(6.42)的一个圆盘之中.

(2) 如果(6.42)的n个圆盘中有k个圆盘之并形成一个连通域,而且与其余$n-k$个圆盘不相交,那么该连通域正好包含k个A的特征值.

证 设λ是A的特征值,$x=(x_1,\cdots,x_n)^{\mathrm{T}}$是相应的特征向量,$Ax=\lambda x$.而且,设$x$规范化成

$$\max_{1\le i\le n}|x_i|=1,\quad x_p=1;\quad |x_j|\le 1,\quad j\ne p.$$

则$Ax=\lambda x$的第p个方程可以写成

$$\lambda-a_{pp}=\sum_{j=1,j\ne p}^{n}a_{pj}x_j.$$

于是

$$|\lambda-a_{pp}|\le\sum_{j=1,j\ne p}^{n}|a_{pj}||x_j|\le\sum_{j=1,j\ne p}^{n}|a_{pj}|.$$

这表明

$$\lambda\in G_p(A)\subset G(A),$$

(1)得证.

现在证明(2).将 A 表示成

$$A \equiv D + B, \quad D \equiv \operatorname{diag}(a_{11}, \cdots, a_{nn}).$$

考虑

$$A_\varepsilon = D + \varepsilon B, \quad \varepsilon \in [0,1],$$

特别

$$A_0 = D, \quad A_1 = A.$$

利用(6.42)的记号,

$$G_i(A_\varepsilon) \equiv \left\{ z : |z - a_{ii}| \le \varepsilon \sum_{j=1, j\ne i}^{n} |a_{ij}| \right\}, \quad i = 1, \cdots, n; \quad z \in \mathbb{C}.$$

A_ε 的所有特征值包含在

$$G(A_\varepsilon) \equiv \bigcup_{i=1}^{n} G_i(A_\varepsilon)$$

之中.

为了方便,设(6.42)中前 k 个圆盘之并

$$\hat{G}(A) \equiv \bigcup_{i=1}^{k} G_i(A)$$

是一个连通域,而且

$$\hat{G}(A) \cap G_i(A) = \varnothing, \quad i = k+1, \cdots, n,$$

其中 $\varnothing$ 表示空集.注意到 $A_1 = A$, A_ε 的前 k 个圆盘之并

$$\hat{G}(A_\varepsilon) \equiv \bigcup_{i=1}^{k} G_i(A_\varepsilon) \subset \hat{G}(A_1) = \hat{G}(A), \quad \varepsilon \in [0,1].$$

自然, $\hat{G}(A_\varepsilon)$ 本身不一定对所有 $\varepsilon \in [0,1]$ 来说都是连通域.但是,必有

$$\hat{G}(A_\varepsilon) \cap G_i(A) = \varnothing, \quad i = k+1, \cdots, n; \varepsilon \in [0,1]. \quad (6.44)$$

用 $\lambda_1(\varepsilon), \cdots, \lambda_n(\varepsilon)$ 表示 A_ε 的特征值,于是

$$\lambda_i(0) = a_{ii}, \quad i = 1, \cdots, n,$$

它们是 $D \equiv A_0$ 的特征值.而

$$\lambda_1(1), \cdots, \lambda_n(1)$$

就是 $A \equiv A_1$ 的特征值.由此,以及特征值是矩阵的元素的连续函数,推出当 $\varepsilon \in [0,1]$ 且足够小时,

$$\lambda_i(\varepsilon) \in G_i(A_\varepsilon) \subset G_i(A), \quad i = 1, \cdots, n. \tag{6.45}$$

特别,根据(6.44),对于 $i = 1, \cdots, k$,

$$\{\lambda_i(\varepsilon) : 0 \le \varepsilon \le 1\}$$

是 $\hat{G}(A)$ 中的连续曲线.这样, 对于每个 $\varepsilon \in [0,1]$,至少有 A_ε 的 k 个特征值包含在 $\hat{G}(A_\varepsilon)$ 中.然而,从(6.44)和(6.45)及特征值连续依赖 ε 得知,$\hat{G}(A_\varepsilon)$ 不可能包含多于 k 个 A_ε 的特征值.因此 $\hat{G}(A)$ 正好包含 A 的 k 个特征值(按重数计). □

(6.43)中的 $G(A)$ 称为 A 的(关于行的)**Gerschgorin 域**; (6.42)中的每个 $G_i(A)$ 称为 A 的 **Gerschgorin 圆盘**,它们的边界称为 A 的 **Gerschgorin 圆**.

Gerschgorin 定理给出了估计矩阵特征值位置的简单方法.

1.6.24 推论 设 $A = [a_{ij}] \in \mathbb{C}^{n \times n}$,则

(1) A 的所有特征值包含在如下 n 个圆盘的并集之中:

$$G(A^{\mathrm{T}}) \equiv \bigcup_{j=1}^{n} \left\{ z : |z - a_{jj}| \le \sum_{i=1, i \ne j}^{n} |a_{ij}| \right\}, \quad z \in \mathbb{C}. \tag{6.46}$$

(2) 如果 $G(A^{\mathrm{T}})$ 中有 k 个圆盘之并形成一个连通域,而且与其余 $n-k$ 个圆盘不相交,那么该连通域正好包含 k 个 A 的特征值.

证 将 **1.6.23** 的(1)和(2)分别应用于 A^{T},并注意到 A^{T} 和 A 有相同的特征值. □

1.6.25 推论 设 $A = [a_{ij}] \in \mathbb{C}^{n \times n}$, $p_1, \cdots, p_n$ 是正实数,则 A 的所有

特征值包含在如下域中:

$$G(D^{-1}AD)\equiv\bigcup_{i=1}^{n}\left\{z:|z-a_{ii}|\le\frac{1}{p_i}\sum_{j=1,j\ne i}^{n}p_j|a_{ij}|\right\},\ z\in\mathbb{C},\tag{6.47}$$

也包含在如下域中:

$$G((D^{-1}AD)^T)\equiv\bigcup_{j=1}^{n}\left\{z:|z-a_{jj}|\le p_j\sum_{i=1,i\ne j}^{n}\frac{1}{p_i}|a_{ij}|\right\},\ z\in\mathbb{C},\tag{6.48}$$

其中 $D\equiv\mathrm{diag}(p_1,\cdots,p_n)$.

证 将 **1.6.23** 和 **1.6.24** 应用于

$$D^{-1}AD=[p_ja_{ij}/p_i],$$

并注意矩阵的相似变换不改变其特征值. □

1.6.26 引理 设 $A=[a_{ij}]\in\mathbb{C}^{n\times n}$, λ 是位于 $G(A)$ 边界上的 A 的特征值. 设

$$Ax=\lambda x,\quad x=(x_1,\cdots,x_n)^{\mathrm{T}}\ne 0,$$

p 是一个指标,使得

$$|x_p|=\max_{1\le i\le n}|x_i|\ne 0,$$

则

(1) 如果 k 是使得 $|x_k|=|x_p|$ 的任一指标,那么成立

$$|\lambda-a_{kk}|=\sum_{j=1,j\ne k}^{n}|a_{kj}|.\tag{6.49}$$

也就是说, A 的第 k 个 Gerschgorin 圆通过 λ.

(2) 如果对于某 k 有

$$|x_k|=|x_p|,\quad 1\le k\le n,$$

那么对于凡有 $a_{kj}\ne 0$ 的 j, 必有

$$|x_j|=|x_p|.$$

证 依假设,λ 满足

$$(\lambda - a_{ii})x_i = \sum_{j=1, j\neq i}^{n} a_{ij}x_j, \quad i=1,\cdots,n.$$

因此

$$\begin{aligned} |\lambda - a_{ii}||x_i| &= \sum_{j=1,j\neq i}^{n} |a_{ij}||x_j| \\ &\leq \left(\sum_{j=1,j\neq i}^{n} |a_{ij}|\right)|x_p|, \quad i=1,\cdots,n. \end{aligned} \tag{6.50}$$

于是,如果 k 是使得 $|x_k|=|x_p|$ 的任一指标,那么,一方面

$$|\lambda - a_{kk}| \leq \sum_{j=1,j\neq k}^{n} |a_{kj}|.$$

另一方面,因 λ 位于 $G(A)$ 的边界,故必有

$$|\lambda - a_{ii}| \geq \sum_{j=1,j\neq i}^{n} |a_{ij}|, \quad i=1,\cdots,n.$$

推出(6.49)成立,而且由(6.50)有

$$|\lambda - a_{kk}||x_k| = \sum_{j=1,j\neq k}^{n} |a_{kj}||x_j| = \sum_{j=1,j\neq k}^{n} |a_{kj}||x_k|.$$

由此,得

$$\sum_{j=1,j\neq k}^{n} |a_{kj}|(|x_k| - |x_j|) = 0.$$

注意到

$$|x_k| - |x_j| \geq 0, \quad j=1,\cdots,n,$$

故对于凡有 $a_{kj}\neq 0$ 的 j,必有 $|x_j|=|x_p|$. □

1.6.27 定理 设 $A=[a_{ij}]\in\mathbb{C}^{n\times n}$,$\lambda$ 是 A 的特征值而且是 $G(A)$ 的边界点,且 A 的所有元素不为零.则

(1) A 的每个 Gerschgorin 圆通过 λ.

(2) 如果

$$Ax = \lambda x, \quad x = (x_1, \cdots, x_n)^{\mathrm{T}} \neq 0,$$

那么$|x_1| = \cdots = |x_n|$.

证 设$|x_p| = \max\limits_{1 \le i \le n} |x_i|$,从**1.6.26**推证得知(6.51)和(6.52)对$k = p$成立.并且注意到现在A的所有元素不为零,立即推出结论(2).然后再利用**1.6.26**的(1),即得本定理的结论(1). □

1.7 Euclid 结构

1.7.1 引言 先来考察一下Euclid空间基本结构的具体背景.

在实数域上的n维Euclid空间中,选取一点O作为原点;然后,引进一个笛卡儿坐标系,并且将空间中任一**向量**x的笛卡儿坐标记作

$$(x_1, \cdots, x_n)^{\mathrm{T}}.$$

x的**长度**定义为它到O的**距离**,记作$\|x\|$.

反复使用Pythagoras定理(即勾股定理),可以将x的**长度**(也称**范数**)通过其笛卡儿坐标表示为

$$\|x\| = \sqrt{x_1^2 + \cdots + x_n^2}. \tag{7.1}$$

两个向量$x = (x_1, \cdots, x_n)^{\mathrm{T}}$和$y = (y_1, \cdots, y_n)^{\mathrm{T}}$的**内积**(也称**数积**),记作$(x, y)$,定义为

$$(x, y) \equiv \sum_{i=1}^{n} x_i y_i. \tag{7.2}$$

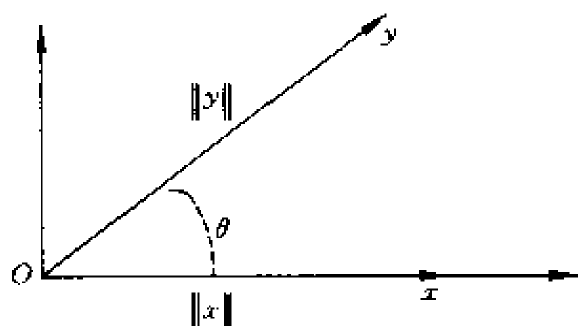
以上两个概念是有联系的:一方面,向量的长度可以表示为

$$\|x\| = (x,x)^{1/2}. \tag{7.3}$$

另一方面,两个向量的内积能够表示为

$$(x,y) = \left\|\frac{x+y}{2}\right\|^2 - \left\|\frac{x-y}{2}\right\|^2. \tag{7.4}$$

选取特殊的坐标轴,第1根轴平行于 x,第2根轴使得 y 被包含在由头两根轴生成的平面上,如此便可以揭示内积 (x,y) 的几何意义见图.



在这种坐标系中,向量 x 和 y 的坐标表示是

$$x = \left(\|x\|, 0, \cdots, 0\right)^{\mathrm{T}}$$

和

$$y = \left(\|y\|\cos\theta, \|y\|\sin\theta, 0, \cdots, 0\right)^{\mathrm{T}},$$

θ 是 x 和 y 之间的夹角.所以

$$(x,y) = \|x\|\|y\|\cos\theta, \tag{7.5}$$

直观上看,只要向量 x 给定了,其长度就是确定的,因此尽管 x 的坐标会因笛卡儿坐标系的不同选取而不同,但是在任何坐标系中公式(7.1)的值应是相同的.

因此从公式(7.4)推出任何两个给定向量的内积在所有笛卡儿坐标系中有相同的值.

下面给出一个抽象的,也就是公理系统的 Euclid 空间定义.

1.7.2 定义 设 X 是实数域上的线性空间.

所谓在 X 中赋予了**实 Euclid 结构**就是提供了一个称为**内积**的两个向量自变量的实值函数,并记之为 $(\bullet,\bullet)$,它必须满足如下三条性质:

(1) $(\bullet,\bullet)$ 是**双线性**函数,也就是说,它对每个向量自变量来说,当另一个保持固定时是线性函数.

(2) 它具有**对称性**,即

$$(x,y)=(y,x), \quad \forall x,y\in X. \tag{7.6}$$

(3) 它具有**正定性**,即

$$(x,x)>0, \quad \forall x\in X, x\neq 0. \tag{7.7}$$

赋予实 Euclid 结构的线性空间称为**实 Euclid 空间**.

内积(7.2)是常用的 Euclid 结构,显然满足上述三条公理.

以下,将看到上述三条简单的公理蕴涵着全部 Euclid 几何.

1.7.3 定义 设 $(\bullet,\bullet)$ 是 Euclid 空间的内积,是由 $(\bullet,\bullet)$ 确定的单个向量自变量的非负函数:

$$\|x\|\equiv(x,x)^{1/2}, \quad \forall x\in X. \tag{7.8}$$

$\|\bullet\|$ 称为 **Euclid 长度**或 **Euclid 范数**.

根据这个定义,从实 Euclid 空间的内积双线性和对称性,得

$$\|x+y\|^2=\|x\|^2+2(x,y)+\|y\|^2. \tag{7.9}$$

由此恒等式立即推出对任一内积成立(7.4).公式(7.4)称为**平行四边形定律**.

1.7.4 定义 在具有 Euclid 范数的线性空间中,任何两个向量的 x 和

y 的**距离**定义为 $\|x-y\|$.

1.7.5 Schwarz 不等式　设 X 是 Euclid 空间.则成立

$$|(x,y)| \le \|x\|\|y\|, \quad x,y \in X. \tag{7.10}$$

证　当 $y=0$ 时(7.10)中的不等式显然成立.

现在任意取定 $x,y \in X$, $y \neq 0$.考虑实变量函数 q,

$$q(t) \equiv \|x+ty\|^2, \quad \forall t \in \mathbb{R}.$$

利用(7.8)和(7.9)有

$$q(t) = \|x\|^2 + 2t(x,y) + t^2\|y\|^2, \tag{7.11}$$

由于 $y \neq 0$,在(7.11)中置

$$t = -(x,y)/\|y\|^2,$$

则从 q 有非负性得到

$$\|x\|^2 - \frac{(x,y)^2}{\|y\|^2} \ge 0,$$

此不等式等价于(7.10)中的不等式.　□

对于具体内积(7.2),不等式(7.10)可直接从(7.5)推得.

不难证明(留作练习):

$$\|x\| = \max_{\|y\|=1}(x,y). \tag{7.12}$$

1.7.6 三角不等式　设 X 是 Euclid 空间.则成立

$$\|x+y\| \le \|x\| + \|y\|, \quad \forall x,y \in X. \tag{7.13}$$

证　在(7.9)右边的中间一项利用估计式(7.10).　□

1.7.7 定义　Euclid 空间两个向量 x 和 y 称为**正交(垂直)**,记作 $x \perp y$,如果

$$(x,y)=0.$$

从(7.9)推出 **Pythagoras 定理**: 若 $x \perp y$,则

$$\|x+y\|^2=\|x\|^2+\|y\|^2. \tag{7.14}$$

1.7.8 定义 设 X 是 n 维 Euclid 空间,而且

$$x^{(1)},\cdots,x^{(n)}$$

是 X 的一组基.此基称为关于给定的 Euclid 结构是**正交**的,如果

$$\left(x^{(i)},x^{(j)}\right)=\begin{cases}0, & i\neq j,\\ 1, & i=j.\end{cases} \tag{7.15}$$

1.7.9 Gram-Schmidt 定理 设 $y^{(1)},\cdots,y^{(n)}$ 是 n 维 Euclid 空间 X 的任一组基,则存在相关的基 $x^{(1)},\cdots,x^{(n)}$ 具有如下性质:

(1) $x^{(1)},\cdots,x^{(n)}$ 是 X 的一组正交基.

(2) $x^{(k)}$ 是 $y^{(1)},\cdots,y^{(k)}$ 的线性组合, $k=1,\cdots,n$.

证 采用递推过程.

假定 $x^{(1)},\cdots,x^{(k-1)}$ 已经构造完毕,置

$$x^{(k)}=c\left(y^{(k)}-\sum_{j=1}^{k-1}c_j x^{(j)}\right).$$

因 $x^{(1)},\cdots,x^{(k-1)}$ 已经正交,容易看出要使如上形式的 $x^{(k)}$ 与它们正交只须选取

$$c_i=\left(y^{(k)},x^{(i)}\right),\quad i=1,\cdots,k-1.$$

然后选取 c 使得 $\left\|x^{(k)}\right\|=1$. □

这一定理保证 n 维 Euclid 空间存在很多正交基.

对于 n 维 Euclid 空间 X 中取定的正交基 $x^{(1)},\cdots,x^{(n)}$,利用正

交关系(7.15),立即得到任一$x \in X$有如下表示式:

$$x=\sum_{i=1}^{n} a_i x^{(i)}; \quad a_i=\left(x, x^{(i)}\right), \quad i=1,\cdots,n . \qquad (7.16)$$

现设

$$x, y \in X, \quad x=\sum_{i=1}^{n} a_i x^{(i)}, \quad y=\sum_{i=1}^{n} b_i x^{(i)},$$

则

$$(x, y)=\sum_{i=1}^{n} \sum_{j=1}^{n} a_i b_j\left(x^{(i)}, x^{(j)}\right)=\sum_{i=1}^{n} a_i b_i . \qquad (7.17)$$

特别,取$y=x$时,得到

$$\|x\|^2=\sum_{i=1}^{n} a_i^2 . \qquad (7.18)$$

比较(7.17)和(7.2),可以得出这样的结论:由

$$x \to \left(a_1, \cdots, a_n\right)^{\mathrm{T}} \qquad (7.19)$$

确定的是从n维 Euclid 空间X到$\mathbb{R}^n$的同构映射,其中a_i是x关于X的一组正交基的第i个分量,这个同构映射将X中给定的内积转换为$\mathbb{R}^n$中的标准内积(7.2).

1.7.10 定义 设X是 Euclid 空间,序列$\left\{x^{(i)}\right\} \in X$.如果存在向量$x \in X$,使得

$$\lim_{i \to \infty}\left\|x^{(i)}-x\right\|=0,$$

则称$\left\{x^{(i)}\right\}$**收敛于极限**x,记作$\lim\limits_{i \to \infty} x^{(i)}=x$.

1.7.11 定义 设X是Euclid空间,序列$\left\{x^{(i)}\right\} \in X$称为**Cauchy序列**,如果当$i \to \infty$与$j \to \infty$时

$$\left\|x^{(i)}-x^{(j)}\right\| \to 0 .$$

可以证明:在有限维Euclid空间中每个Cauchy序列收敛于一极限.这一性质称为有限维Euclid空间的**完备性**.

1.7.12 定义 设 X 是 Euclid 空间,序列 $\{x^{(i)}\}\in X$ 称为**有界**,如果存在正数 R,使得

$$\|x^{(i)}\|\le R,\quad i=1,2,\cdots.$$

可以证明: 在有限维 Euclid 空间中每个有界序列包含收敛的子序列. 这一性质称为有限维 Euclid 空间的**局部紧性**.

1.7.13 定理 l 是有限维Euclid空间 X 上线性函数的充分必要条件为 l 形如

$$l(x)\equiv(x,y),\quad \forall x\in X, \tag{7.20}$$

y 是 X 的某个元素.

推论:映射 $y\mapsto l$ 是 X 与其对偶 X' 的同构映射.

证 充分性. l 形如(7.20).因为内积是双线性的,且 y 是固定的,所以 l 是线性函数.

必要性. l 是线性函数.在 X 中取一组正交基 $x^{(1)},\cdots,x^{(n)}$,记

$$b_i\equiv l(x^{(i)}),\quad i=1,\cdots,n.$$

令

$$y\equiv\sum_{i=1}^{n}b_i x^{(i)},$$

从正交性推出

$$(x^{(i)},y)=b_i,\quad i=1,\cdots,n.$$

这表明(7.20)对

$$x=x^{(i)},\quad i=1,\cdots,n$$

成立;然而,假若两个线性函数对于构成一组基的每个向量处都取相同的值,那么这两个线性函数是恒等的. □

1.7.14 定义 设 X 是有限维 Euclid 空间，Y 是 X 的线性子空间。Y 的**正交补**记作 $Y^{\perp}$，它是由 X 中正交于 Y 的所有向量组成的，即

$$Y^{\perp} \equiv \{z \in X : (y,z) = 0, \forall y \in Y\}.$$

在 **1.2.6** 中曾用 $Y^{\perp}$ 表示在 Y 上等于零的线性函数的集合。当通过 **1.7.13** 把 X 的对偶 X' 与 X 视为同一时，上面引入的 $Y^{\perp}$ 和 **1.2.6** 中的 $Y^{\perp}$ 是一致的。特别，$Y^{\perp}$ 是 X 的线性子空间。

1.7.15 定理 设 X 是有限维 Euclid 空间，Y 是 X 的线性子空间。则 X 是 Y 与其正交补 $Y^{\perp}$ 的直和：

$$X = Y \oplus Y^{\perp}.$$

等价地说，每个 $x \in X$ 可以唯一地分解为

$$x = y + y^{\perp}, \quad y \in Y, \quad y^{\perp} \in Y^{\perp}. \tag{7.21}$$

证 首先证明形如(7.21)的分解是唯一的。假定还有

$$x = z + z^{\perp}, \quad z \in Y, \quad z^{\perp} \in Y^{\perp}.$$

将其与(7.21)相减，得

$$y - z = z^{\perp} - y^{\perp}.$$

这说明 $y - z$ 同时属于 Y 和 $Y^{\perp}$，从而它与自己正交：

$$0 = (y - z, z^{\perp} - y^{\perp}) = (y - z, y - z) = \|y - z\|^2.$$

由范数的正定性，$y - z = 0$，即 $z = y$。

为了证明形如(7.21)的分解总是可能的，构造 X 的一组正交基 $x^{(1)}, \cdots, x^{(n)}$，其前 k 个成员在 Y 中，则其余成员必在 $Y^{\perp}$ 中。构造如此正交基可以这样来实现：按照 **1.7.9** 中的正交化步骤，先产生 Y 中一组正交基，然后将其扩充成 X 的一组基，并且把新扩充进来的成员加以正交化。于是

$$x = \sum_{i=1}^{n} a_i x^{(i)} \equiv y + y^{\perp}; \quad y \equiv \sum_{i=1}^{k} a_i x^{(i)}, \quad y^{\perp} \equiv \sum_{i=k+1}^{n} a_i x^{(i)}.$$

显然，$y \in Y, y^{\perp} \in Y^{\perp}$. □

1.7.16 定义 设 X 是有限维 Euclid 空间，Y 是 X 的线性子空间，$x \in X$ 有分解式

$$x = y + y^{\perp}, \quad y \in Y, \quad y^{\perp} \in Y^{\perp},$$

则称 y 为 x 在 Y 中的**正交投影**，记作

$$y = P_Y x. \tag{7.22}$$

如此映射 $P_Y : X \to Y$ 称为**正交投影映射**.

1.7.17 定理 设 X 是有限维 Euclid 空间，Y 是 X 的线性子空间，P_Y 是正交投影映射，则

(1) P_Y 是线性映射.

(2) P_Y 具有等幂性：$P_Y^2 = P_Y$.

证 设任意的 $u, v \in X$ 有如下分解：

$$u = y + y^{\perp}, v = z + z^{\perp}; \quad y, z \in Y, \quad y^{\perp}, z^{\perp} \in Y^{\perp}.$$

由此得 $u + v$ 的分解

$$u + v = (y + z) + \left(y^{\perp} + z^{\perp}\right) \quad y + z \in Y, \quad y^{\perp} + z^{\perp} \in Y^{\perp}.$$

依 **1.7.16**

$$P_Y(u + v) = y + z = P_Y u + P_Y v.$$

类似地

$$P_Y(\alpha u) = \alpha y = \alpha P_Y u,$$

所以 P_Y 是线性映射.

由于 $y = P_Y u \in Y$ 无须再分解，从而依 **1.7.16**，有

$$P_Y^2 u = P_Y\left(P_Y u\right) = P_Y y = y = P_Y u.$$

因此 $P_Y^2 = P_Y$. □

1.7.18 定理 设 X 是有限维 Euclid 空间，Y 是 X 的线性子空间，向量 $x \in X$. 则 Y 中与 x 之间 Euclid 距离最近的向量是 $P_Y x$.

证 x 可以写成

$$x = P_Y x + y^{\perp}, \quad y^{\perp} \in Y^{\perp}.$$

于是,$\forall z \in Y$,

$$x - z = (P_Y x - z) + y^{\perp}; \quad P_Y x - z \in Y, \quad y^{\perp} \in Y^{\perp}.$$

根据 Pythagoras 定理(7.14),

$$\|x - z\|^2 = \|P_Y x - z\|^2 + \|y^{\perp}\|^2, \quad \forall z \in Y.$$

显然,当 $z = P_Y x$ 时 $\|x - z\|$ 达最小. □

1.7.19 定义 设 X 是有限维 Euclid 空间;$A \in \mathscr{L}(X, X)$,即 A 是 X 到 X 的线性映射.对于给定的 $x \in X$,令

$$l(x) \equiv (Ax, y), \quad \forall x \in X.$$

注意到内积是双线性的,l 是 X 上的线性函数.依 **1.7.13**,存在某 $z \in X$,使得

$$l(x) = (x, z), \quad \forall x \in X,$$

因此

$$(Ax, y) = (x, z), \quad \forall x \in X.$$

此式表明了 z 和 y 的依赖关系,而且这种依赖是线性的.将其表示为 $z \equiv A^* y$,于是 $A^* \in \mathscr{L}(X, X)$,而且

$$(Ax, y) = (x, A^* y), \quad \forall x, y \in X, \tag{7.23}$$

A^* 称为 A 的**伴随映射**.

平行地,引出矩阵的相应概念.设 $\dim X = n$,$A \in \mathscr{L}(X, X)$ 对应的矩阵仍记作 A,其伴随映射 $A^* \in \mathscr{L}(X, X)$ 对应的矩阵仍记作 A^*,此时 A 和 A^* 均为 $n \times n$ 矩阵,A^* 称为 A 的**伴随矩阵**.

注意这里定义的伴随矩阵与 **1.5.13** 中定义的伴随矩阵两者间的区别.

容易证明:对于实数域上的有限维 Euclid 空间 X 来说,

$A \in \mathscr{L}(X,X)$ 的伴随映射 A^* 即为 A 的转置，$A^* = A'$ (见 **1.3.8**). 当 $X = \mathbb{R}^n$ 且取标准 Euclid 结构(7.2)时，$n \times n$ 矩阵 A 的伴随矩阵 A^* 就是 A 的转置矩阵，$A^* = A^{\mathrm{T}}$ (见 **1.4.3**).

容易证明:若 P_Y 是如 **1.7.16** 定义的正交投影映射,则

$$P_Y^* = P_Y. \tag{7.24}$$

1.7.20 定理 伴随矩阵有如下性质:

(1) $(A+B)^* = A^* + B^*$.

(2) $(AB)^* = B^* A^*$.

(3) $(A^{-1})^* = (A^*)^{-1}$.

(4) $(A^*)^* = A$.

(5) $\mathscr{N}_{A^*A} = \mathscr{N}_A$, $\mathscr{R}_{A^*A} = \mathscr{R}_{A^*}$.

(6) 若 $A^*AB = A^*AC$,则 $AB = AC$.

(7) 若 $BA^*A = CA^*A$,则 $BA^* = CA^*$.

证 (1) 从(7.23)直接推出.

(2) 分两步走:

$$(ABx, y) = (Bx, A^* y) = (x, B^* A^* y).$$

(3) 是(2)的一个推论.

(4) 从(7.23)与内积的对称性推出.

(5) 显然,$\mathscr{N}_{A^*A} \supset \mathscr{N}_A$.反之,若 $y \in \mathscr{N}_{A^*A}$,即 $A^*Ay = 0$,则

$$\|Ay\|^2 = (Ay, Ay) = (y, A^*Ay) = 0,$$

从而 $Ay = 0$,即 $y \in \mathscr{N}_A$;于是 $\mathscr{N}_{A^*A} \subset \mathscr{N}_A$,因此 $\mathscr{N}_{A^*A} = \mathscr{N}_A$.

类似地,显然,$\mathscr{R}_{A^*A} \subset \mathscr{R}_{A^*}$.反之,若 $y = A^*x \in \mathscr{R}_{A^*}$,注意到

$$\begin{aligned} \operatorname{rank} A^*A &\le \operatorname{rank}\begin{bmatrix} A^*A & A^*x \end{bmatrix} \\ &\le \operatorname{rank} A^* = \operatorname{rank} A = \operatorname{rank} A^*A, \end{aligned}$$

即有$\operatorname{rank}\begin{bmatrix} A^*A & A^*x \end{bmatrix} = \operatorname{rank}A^*A$,于是方程组

$$A^*Az = A^*x(\equiv y)$$

可解,表明$y \in \mathscr{R}_{A^*A}$.因此$\mathscr{R}_{A^*A} = \mathscr{R}_{A^*}$.

(6) 因为$A^*A(B-C)=0$,从(5)推出$A(B-C)=0$.

(7) 根据假定,

$$A^*AB^* = (BA^*A)^* = (CA^*A)^* = A^*AC^*$$

从(6)得$AB^* = AC^*$,等价于$BA^* = CA^*$. □

1.7.21 定义 设X是有限维Euclid空间,$A \in \mathscr{L}(X,X)$,

$$\|A\| \equiv \max_{x\in X, x\neq 0} \frac{\|Ax\|}{\|x\|}, \tag{7.25}$$

其中右边分子分母中所取的范数是X中的向量范数.$\|A\|$称为线性映射A的**范数**.

平行地,引出矩阵的相应概念.设$\dim X = n$,而且$A \in \mathscr{L}(X,X)$对应的$n\times n$矩阵仍记作A.此时,由(7.25)定义的$\|A\|$称为矩阵A的**范数**.

不难证明,对于有限维Euclid空间X来说,(7.25)定义中的最大值是存在的.这样,有一种度量向量长度的方式,就可以由(7.25)给出一种相应的度量线性映射$A \in \mathscr{L}(X,X)$"大小"的方式;这就是空间$\mathscr{L}(X,X)$的范数.

从(7.25)推出

$$\|Ax\| \le \|A\|\|x\|, \quad \forall x \in X. \tag{7.26}$$

1.7.22 定理 设X是有限维Euclid空间.空间$\mathscr{L}(X,X)$的范数有以下基本性质:

(1) $\|\alpha A\| = |\alpha|\|A\|, \quad \forall \alpha \in \mathbb{R}, A \in \mathscr{L}(X,X)$.

(2) $\|A+B\| \le \|A\| + \|B\|, \quad \forall A,B \in \mathscr{L}(X,X)$.

(3) $\|AB\| \le \|A\|\|B\|, \quad \forall A,B \in \mathscr{L}(X,X)$.

证明留作练习.

1.7.23 定理 设 X 是有限维 Euclid 空间, $A \in \mathscr{L}(X,X)$,则

(1) $\|Ax\|$ 在单位球面 $\|x\| = 1$ 上有界且到达最大值,并且

$$\|A\| = \max_{\|x\|=1} \|Ax\|. \tag{7.27}$$

(2) 成立

$$\|A\| = \max_{\|x\|=\|y\|=1} (Ax, y). \tag{7.28}$$

(3) 成立

$$\|A^*\| = \|A\|. \tag{7.29}$$

(4) A 是连续的,这是指:

$$\lim_{k\to\infty} x^{(k)} = x \Rightarrow \lim_{k\to\infty} Ax^{(k)} = Ax.$$

证明留作练习.

1.7.24 定义 设 X 是有限维 Euclid 空间, M 是 $X \to X$ 的映射且保持任何**点对**的距离,即满足

$$\|Mx - My\| = \|x - y\|, \quad \forall x, y \in X, \tag{7.30}$$

则称 M 为 X 的**等距映射**.

由定义即知:两个等距映射的复合是等距映射.

等距映射的一个初等例子是**平移**:

$$Mx \equiv x + a, \quad \forall x \in X, \tag{7.31}$$

$a \in X$ 是某固定向量.

可以用一个平移和一个将零映成零的等距映射复合成一个等

距映射.反之,任一等距映射均是一个将零映成零的等距映射和一个平移的复合.

1.7.25 定理 设 X 是有限维 Euclid 空间, M 是 X 的将零映成零的等距映射:

$$M0=0, \tag{7.32}$$

则

(1) M 是线性映射.

(2) 成立

$$M^*M=I. \tag{7.33}$$

反之,如果(7.33)成立,那么 M 是等距映射.

(3) M 是可逆的,且其逆是等距映射.

(4) $\det M=\pm1$,这里的 M 理解为是等距映射 M 对应的矩阵.几何意义: M 不仅是保持距离还是保持体积的映射.

证 (1)从(7.30)取 $y=0$ 及从(7.32)得

$$\|Mx\|=\|x\|,\quad \forall x,y\in X. \tag{7.34}$$

另一方面,由(7.30),

$$\begin{aligned}&(Mx-My,Mx-My)\\&\qquad=\|Mx-My\|^2=\|x-y\|^2=(x-y,x-y),\end{aligned}$$

将两端的内积展开并利用(7.34),得

$$(Mx,My)=(x,y),\quad \forall x,y\in X. \tag{7.35}$$

这就是说, M 是保持内积的.

仍利用内积展开,有

$$\begin{aligned}\|Mz-Mx-My\|^2=&\|Mx\|^2+\|My\|^2+\|Mz\|^2\\&-2(Mz,Mx)-2(Mz,My)+2(Mx,My)\end{aligned}$$

和

$$\|z-x-y\|^2=\|z\|^2+\|x\|^2+\|y\|^2-2(z,x)-2(z,y)+2(x,y).$$

利用(7.34)和(7.35),得

$$\|Mz-Mx-My\|^2=\|z-x-y\|^2.$$

由此,取 $z = x + y$,便有

$$M(x+y) \equiv Mz = Mx + My.$$

类似地,可从

$$\|Mz - \alpha Mx\|^2 = \|Mz\|^2 + \alpha^2\|Mx\|^2 - 2\alpha(Mz, Mx)$$

和

$$\|z - \alpha x\|^2 = \|z\|^2 + \alpha^2\|x\|^2 - 2\alpha(z, x)$$

推出

$$M(\alpha x) = \alpha Mx.$$

(1)得证.

(2).根据(7.23)和(7.35),

$$(x, M^*My) = (Mx, My) = (x, y), \quad \forall x, y \in X.$$

于是

$$(x, M^*My - y) = 0, \quad \forall x, y \in X.$$

特别,取 $x = M^*My - y$,推出 $M^*My - y$ 与自己正交,因此由范数的正定性得

$$M^*My - y = 0, \quad \forall y \in X.$$

这等价于 $M^*M = I$.

(3) 从(7.34)推出 M 的零空间仅由零向量组成;故由 **1.3.4** 的(2)知 M 是可逆的.再从(7.34)和 M 的线性可推出 M^{-1} 是等距的.(3)得证.

(4) 依 **1.5.13** 中的(5.33), $\det M^* = \det M$;再依 **1.5.8**,

$$(\det M)^2 = \det M^* \det M = \det(M^*M) = \det I = 1,$$

即得 $\det M = \pm 1$. □

1.7.26 定义 将$\mathbb{R}$"等距地映射成$\mathbb{R}$"的(线性映射对应的)矩阵称为**正交矩阵**(orthogonal matrix).

根据 **1.7.25**,正交矩阵的行列式等于 $+1$ 或 -1.

同阶的正交矩阵在矩阵乘法下构成一个群.行列式等于$+1$的正交矩阵构成一个子群,称为**特殊正交群**.

1.7.27 定理　下列三个条件等价:

(1) P是正交矩阵.

(2) P的各列为两两正交的单位向量.

(3) P的各行为两两正交的单位向量.

证明留作练习.

1.7.28 定义　在复数域上的线性空间X中赋予**复 Euclid 结构**就是提供一个称为**内积**的两个向量自变量的复值函数,并记作$(\cdot,\cdot)$,它满足如下性质:

(1) 当y固定时,(x,y)是x的**线性**函数.

(2) 它具有**斜对称性**,即

$$(x,y)=\overline{(y,x)},\quad \forall x,y\in X, \tag{7.36}$$

其中$\overline{(y,x)}$表示取(y,x)的**复共轭**.(7.36)蕴涵着$\forall x\in X$,(x,x)是实数.

(3) 它具有**正定性**,即

$$(x,x)>0,\quad \forall x\in X, x\neq 0. \tag{7.37}$$

赋予了复 Euclid 结构的线性空间称为**复 Euclid 空间**,也称为**酉空间**.

复 Euclid 空间的理论和实 Euclid 空间的理论类似,但需作少许必要的修改.为了避免理论叙述的重复,仅指出需作细微修改的基本之处.

从复 Euclid 结构的性质(1)和(2)推出,当x固定时,(x,y)是y的**斜线性函数**:

$$(x,\alpha y)=\overline{\alpha}(x,y),\quad \forall\alpha\in\mathbb{C},\ \forall x,y\in X. \tag{7.38}$$

对于复的情形,**标准 Euclid 结构**(7.2)改变成为

$$(x, y) \equiv \sum_{i=1}^{n} x_i \overline{y_i}. \tag{7.39}$$

范数定义 **1.7.3** 适用于复 Euclid 空间.

平行四边形定律(7.9)改变成为

$$\|x + y\|^2 = \|x\|^2 + 2\operatorname{Re}(x, y) + \|y\|^2, \tag{7.40}$$

其中 $\operatorname{Re}\alpha$ 表示复数 α 的实部.

有限维复 Euclid 空间 X 到 X 的线性映射 A 的**伴随映射** A^*,在复 Euclid 结构意义下,仍然由(7.23)定义.

在标准 Euclid 结构(7.39)下,$n \times n$ 矩阵 A 的伴随矩阵 A^* 有区别于 **1.7.19** 中的解释. 记

$$A = [a_{ij}]; \quad (Ax)_i \equiv \sum_{j=1}^{n} a_{ij} x_j, \quad i = 1, \cdots, n.$$

利用(7.39),

$$(Ax, y) \equiv \sum_{i=1}^{n} \left(\sum_{j=1}^{n} a_{ij} x_j \right) \overline{y_i} = \sum_{j=1}^{n} x_j \overline{\left(\sum_{i=1}^{n} \overline{a_{ij}} y_i \right)} \equiv (x, A^* y).$$

因此

$$(A^* y)_j = \sum_{i=1}^{n} \overline{a_{ij}} y_i.$$

这就是说,A^* 是 A 的转置的复共轭:

$$A^* = \overline{A^{\mathrm{T}}} = \overline{A}^{\mathrm{T}} = \left[\overline{a_{ji}} \right].$$

这样,可以引出如下一般矩阵的伴随矩阵的定义.

$A \in \mathbb{C}^{m \times n}$ 的**伴随矩阵**是指 $A^* \in \mathbb{C}^{n \times m}$,

$$A^* \equiv \overline{A^{\mathrm{T}}} = \overline{A}^{\mathrm{T}}. \tag{7.41}$$

通常,A^* 也称为 A 的**共轭转置矩阵**.

对于一般矩阵的伴随矩阵,**1.7.20** 中的基本性质——除(3)要求 A 为方阵外——仍然成立.

等矩映射定义 **1.7.24** 适用于复 Euclid 空间.

1.7.29 定义 有限维复 Euclid 空间 X 到 X 的等距的线性映射称为**酉映射**.

1.7.30 定理 设 X 是复 Euclid 空间,$U \in \mathscr{L}(X,X)$.

(1) U 为酉映射的充分必要条件是 U 满足

$$U^*U = I. \tag{7.42}$$

(2) 若 U 是酉映射,则 U^* 和 U^{-1} 也是酉映射.

(3) 若 U 是酉映射,则 $|\det U| = 1$.

证明可仿照 **1.7.25**,留作练习.

1.8 赋范线性空间

1.8.1 引言 在 **1.7** 中,以内积为基本结构定义 Euclid 空间,引出 Euclid 范数.实际上,也可以以 Euclid 范数为基本结构定义 Euclid 空间,反过来诱导出内积.因此Euclid空间便是一种赋范线性空间.粗略地说,赋范线性空间就是赋予了范数结构的线性空间.注意,一般不能说从范数结构诱导出一种内积.赋范线性空间是一大类应用非常广泛的抽象空间.

在这里,着重介绍范数的一般概念及继 Euclid 范数后的各种常见的范数结构.同时还简介赋范线性空间的对偶空间的范数结构,以及赋范线性空间之间线性映射的范数.

赋范线性空间上的线性函数也称**线性泛函**,赋范线性空间之间的线性映射多称**线性算子**.

1.8.2 定义 设 X 是数域 K 上的线性空间,在 X 中赋予**范数结构**就是提供一个称为**范数**并记作 $\|\bullet\|$ 的单向量自变量的非负函数,它满足如下性质:

(1) **正定性**: $\|x\| \geq 0, \quad \forall x \in X, x \neq 0; \quad \|0\| = 0$.

(2) **次加性**或称**三角不等式**:

$$\|x+y\| \leq \|x\| + \|y\|, \quad \forall x, y \in X.$$

(3) **齐次性**: $\|\alpha x\| = |\alpha| \|x\|, \quad \forall x \in X, \alpha \in K$.

赋予了范数结构的线性空间称为**赋范线性空间**.

同一线性空间 X 可以赋予不同的范数,从而构成不同的赋范线性空间.

利用范数可以定义距离和开球(邻域),从而引出极限等概念.

1.8.3 定义 赋范线性空间 X 中任意两点 x 和 y 的**距离**,记作 $d(x,y)$,

$$d(x,y) \equiv \|x-y\|. \tag{8.1}$$

从范数的次加性容易证明距离成立**三角不等式**:

$$d(x,y) \leq d(x,z) + d(z,y), \quad \forall x, y, z \in X. \tag{8.2}$$

1.8.4 定义 设 X 是赋范线性空间,任意取定一点 $x^{(0)} \in X$ 和一个正数 r,集合

$$B\left(x^{(0)}, r\right) \equiv \left\{x \in X : d(x, x^{(0)}) < r\right\} \tag{8.3}$$

称为中心为 $x^{(0)}$、半径为 r 的**开球**,或 $x^{(0)}$ 的 $r-$**邻域**.

类似地,可以定义**闭球**.

1.8.5 定义 赋范线性空间 X 中的序列 $\left\{x^{(k)}\right\}$ 称为**收敛**至**极限** a,记

作 $\lim_{k\to\infty} x^{(k)} = a$,如果 $a \in X$,而且

$$\lim_{k\to\infty}\left\|x^{(k)} - a\right\| = 0.$$

表达式 $\lim_{k\to\infty}\left\|x^{(k)} - a\right\| = 0$ 可以换用开球的说法:对于任意给定的 $\varepsilon > 0$,存在 $K > 0$,使得

$$x^{(k)} \in B(a,\varepsilon), \quad \forall k > K.$$

1.8.6 定义 赋范线性空间 X 中的集合 S 称为**闭集**,如果 S 中的所有收敛序列的极限均属于 S.

容易证明(留作练习):每个有限维赋范线性空间是闭的.

1.8.7 定义 赋范线性空间 X 中的集合 S 称为**有界集**,如果 S 包含在某个球中,或者说,存在正数 R,使得

$$\|x\| \le R, \quad \forall x \in S.$$

1.8.8 定义 线性空间 $\mathbb{C}^n$ 的常赋范数有:

$\forall x = (x_1,\cdots,x_n)^{\mathrm{T}} \in \mathbb{C}^n$,

$\infty -$**范数**

$$\|x\|_\infty \equiv \max_{1\le i\le n}|x_i|. \tag{8.4}$$

$2-$**范数,即 Euclid 范数**

$$\|x\|_2 \equiv \left(\sum_{i=1}^{n}|x_i|^2\right)^{1/2}. \tag{8.5}$$

$1-$**范数**

$$\|x\|_1 \equiv \sum_{i=1}^{n}|x_i|. \tag{8.6}$$

$p-$**范数**($p \geq 1$ 即 p 是大于等于1的任何实数)

$$\|x\|_p \equiv \left(\sum_{i=1}^{n}|x_i|^p\right)^{1/p}. \tag{8.7}$$

需要证明满足 **1.8.2** 中关于范数的三条性质.

以上所有情形显然满足性质(1)和(3).

容易证明(留作练习): $\infty-$范数和 $1-$范数满足性质(2).

$2-$范数满足性质(2)已在 **1.7.6** 中证明.

$p-$范数以 $\infty-$范数、$2-$范数和 $1-$范数为特殊情形;特别,$\infty-$范数是 $p-$范数当 $p\to\infty$ 时的极限,证明留作练习.

证明 $p-$范数满足性质(2)需要下述不等式.

1.8.9 Hölder 不等式 设 p 和 q 是正数,满足

$$\frac{1}{p}+\frac{1}{q}=1, \tag{8.8}$$

则成立

$$\left|x^T y\right| \leq \|x\|_p \|y\|_q, \quad \forall x, y \in \mathbb{C}^n. \tag{8.9}$$

其中积 $x^{\mathrm{T}}y$ 的定义为:对于 $x=\left(x_1,\cdots,x_n\right)^{\mathrm{T}}$ 和 $y=\left(y_1,\cdots,y_n\right)^{\mathrm{T}}$,

$$x^{\mathrm{T}}y=\sum_{i=1}^{n}x_i\overline{y_i}, \tag{8.10}$$

即 Euclid 空间中的内积 (x,y).

(8.9)中等号成立的充分必要条件是向量

$$\left(|x_1|^p,\cdots,|x_n|^p\right)$$

和

$$\left(|y_1|^q,\cdots,|y_n|^q\right)$$

成比例,而且

$$\operatorname{sgn} x_i = \operatorname{sgn} y_i, \quad i = 1, \cdots, n.$$

Hölder 不等式的证明在许多实分析论著中可以找到,这里从略.当 $p = q = 2$ 时, Hölder 不等式就是 Schwarz 不等式,见 **1.7.5.**

1.8.10 推论 $\forall x = (x_1, \cdots, x_n)^{\mathrm{T}} \in \mathbb{C}^n$,

$$\|x\|_p = \max_{\|y\|_q = 1} x^{\mathrm{T}} y. \tag{8.11}$$

证 不等式(8.9)表明,当 $\|y\|_q = 1$ 时,

$$x^{\mathrm{T}} y \le \|x\|_p.$$

所以为了证明(8.11),只须找到一个向量 $y^{(0)}$, $\|y^{(0)}\|_q = 1$, 使得

$$x^{\mathrm{T}} y^{(0)} = \|x\|_p.$$

在这里,取

$$y^{(0)} = \frac{z}{\|x\|_p^{p/q}}, \quad z = (z_1, \cdots, z_n)^{\mathrm{T}}, \tag{8.12}$$

其中

$$z_i = (\operatorname{sgn} x_i)|x_i|^{p/q}, \quad i = 1, \cdots, n. \tag{8.13}$$

由于

$$\|z\|_q^q = \sum_{i=1}^{n} |z_i|^q = \sum |x_i|^p = \|x\|_p^p,$$

因此

$$\|y^{(0)}\|_q = \frac{\|z\|_q}{\|x\|_p^{p/q}} = \frac{\|x\|_p^{p/q}}{\|x\|_p^{p/q}} = 1. \tag{8.14}$$

而且,从(8.12)及$\frac{1}{p}+\frac{1}{q}=1$,

$$x^{\mathrm{T}}y^{(0)}=\frac{x^{\mathrm{T}}z}{\|x\|_p^{p/q}}=\frac{\sum_{i=1}^{n}|x_i||x_i|^{p/q}}{\|x\|_p^{p/q}}=\frac{\|x\|_p^p}{\|x\|_p^{p/q}}=\|x\|_p. \qquad \square$$

1.8.11 定理 由(8.7)定义的$\|\bullet\|_p$ $(p\ge 1)$是范数.

证 因为$\|\bullet\|_p$显然满足正定性和齐次性,所以只须证明它满足次加性.利用**1.8.10**,对于任意的两个向量$x,y\in\mathbb{C}^n$,

$$\begin{aligned}\|x+y\|_p&=\max_{\|z\|_q=1}(x+y)^{\mathrm{T}}z\\&\le\max_{\|z\|_q=1}x^{\mathrm{T}}z+\max_{\|z\|_q=1}y^{\mathrm{T}}z=\|x\|_p+\|y\|_p. \qquad \square\end{aligned}$$

1.8.12 定义 有限维线性空间X上的两种范数$\|\bullet\|_{\mathrm{I}}$和$\|\bullet\|_{\mathrm{II}}$称为**等价**的,如果存在常数$c$,使得

$$\|x\|_{\mathrm{I}}\le c\|x\|_{\mathrm{II}},\quad \|x\|_{\mathrm{II}}\le c\|x\|_{\mathrm{I}},\quad \forall x\in X. \tag{8.15}$$

1.8.13 定理 有限维线性空间上所有范数是等价的.

证 任何复数域$\mathbb{C}$上的有限维线性空间X均同构于$\mathbb{C}^n$,$n=\dim X$,所以可取$X=\mathbb{C}^n$.

考虑Euclid范数即2−范数:

$$\|x\|_2=\left(\sum_{i=1}^{n}|x_i|^2\right)^{1/2},$$

其中

$$x=(x_1,\cdots,x_n)^{\mathrm{T}}=\sum_{i=1}^{n}x_ie_i\in\mathbb{C}^n,$$

e_i是第i个分量为1其余分量为0的列单位向量.

设$\|\cdot\|$是$\mathbb{C}^n$上的任一其它范数,重复利用次加性和齐次性,可得

$$\|x\| \le \sum_{i=1}^{n} |x_i| \|e_i\|.$$

再利用 Schwarz 不等式(见 **1.7.5**),得

$$\|x\| \le \left(\sum_{i=1}^{n} \|e_i\|^2\right)^{1/2} \left(\sum_{i=1}^{n} |x_i|^2\right)^{1/2} = c\|x\|_2, \qquad (8.16)$$

这里$c \equiv \left(\sum_{i=1}^{n} \|e_i\|^2\right)^{1/2}$.

另一方面,依次加性,

$$\|x\| \le \|x-y\| + \|y\|, \quad \|y\| \le \|x-y\| + \|x\|,$$

有

$$\big|\|x\| - \|y\|\big| \le \|x-y\|.$$

再利用(8.16),

$$\big|\|x\| - \|y\|\big| \le c\|x-y\|_2.$$

这表明$\|\cdot\|$按 Euclid 范数$\|\cdot\|_2$是连续函数.由于有限维空间中单位球面

$$S \equiv \{x : \|x\|_2 = 1\}, \quad x \in \mathbb{C}^n$$

是紧集,所以连续函数$\|\cdot\|$在S上达到其最小值c',而且因

$$\|x\| > 0, \quad \forall x \in S,$$

推出$c' > 0$.因此

$$\|x\| \ge c' > 0, \quad 当\|x\|_2 = 1. \qquad (8.17)$$

注意到$\|\cdot\|$和$\|\cdot\|_2$都是齐次函数,以及当$x \ne 0$时向量$\|x\|_2^{-1} x$满足

$$\big\|\|x\|_2^{-1} x\big\|_2 = 1.$$

从(8.17)推出

$$c'\|x\|_2 \leq \|x\|, \quad \forall x \in \mathbb{C}^n. \tag{8.18}$$

综合(8.16)和(8.18),证明了$\mathbb{C}^n$上的任一种范数与 Euclid 范数等价.这样,鉴于等价概念具有传递性,$\mathbb{C}^n$上的所有范数是等价的. □

1.8.14 定理 设

$$x^{(k)} = \left(x_1^{(k)}, \cdots, x_n^{(k)}\right)^{\mathrm{T}} \in \mathbb{C}^n, \quad k = 1,2,\cdots$$

而且$\boldsymbol{a} = \left(a_1, \cdots, a_n\right)^{\mathrm{T}} \in \mathbb{C}^n$,则序列$\left\{x^{(k)}\right\}$收敛于$a$的充分必要条件是

$$\lim_{k \to \infty} x_i^{(k)} = a_i, \quad i = 1, \cdots, n.$$

证明留作练习.提示:先选定一种范数,比如$\|\bullet\|_\infty$,容易推证对其成立定理的充分必要条件;然后依 **1.8.13**,得知定理的结论与在$\mathbb{C}^n$上范数的选取无关.

1.8.15 定理 设X是范数为$\|\bullet\|$的有限维赋范线性空间,l是定义在X上的线性函数,则存在常数c,使得

$$|l(x)| \leq c\|x\|, \quad \forall x \in X. \tag{8.19}$$

证 依 **1.7.13**,l可以写成

$$l(x) = (x, y),$$

其中y是X的某个元素.根据 Schwarz 不等式(见 **1.7.5**),

$$|l(x)| \leq \|x\|_2 \|y\|_2, \quad \forall x \in X.$$

由此及(8.18),

$$|l(x)| \leq \frac{\|y\|_2}{c'} \|x\|, \quad \forall x \in X.$$

令$c \equiv \|y\|_2 / c'$,即得(8.19). □

1.8.16 定义 设 X' 是范数为 $\|\bullet\|$ 的赋范线性空间 X 的对偶空间，$\|\bullet\|'$ 是定义在 X' 上的函数，

$$\|l\|' \equiv \sup_{x\in X,\|x\|=1} lx, \quad \forall l \in X' \tag{8.20}$$

这里为了对称起见，将值 $l(x)$ 记作 lx，$\|\bullet\|'$ 称为范数 $\|\bullet\|$ 的**对偶范数**.

1.8.15 保证了(8.20)是有意义的，对每个 $l \in X'$，$\|\bullet\|'$ 是有限的数.

1.8.17 定理 对偶范数是对偶空间的范数.

证明留作练习.即验证 $\|\bullet\|'$ 满足范数的三条性质.

利用 **1.8.16** 可以定义 X' 的对偶空间 X'' 的范数:

$$\|x\|'' \equiv \sup_{\|l\|'=1} xl, \quad \forall x \in X''. \tag{8.21}$$

注意，**1.2.5** 表明在同构意义下，$X'' = X$.

1.8.18 定理 赋范线性空间 X 的对偶范数的对偶范数是原始范数，也就是说，

$$\|x\|'' = \|x\|, \quad \forall x \in X. \tag{8.22}$$

此定理的证明从略.

1.8.19 定理 设 X 和 Y 是两个有限维赋范线性空间，范数分别记作 $\|\bullet\|_X$ 和 $\|\bullet\|_Y$.设 $A \in \mathscr{L}(X,Y)$，即 A 是从 X 到 Y 的线性算子，则

存在常数 c,使得

$$\|Ax\|_Y \le c\|x\|_X, \quad \forall x \in X. \tag{8.23}$$

证 设 $\dim X = n$, $x^{(1)},\cdots,x^{(n)}$ 是 X 的一组基.从而,每个 $x \in X$ 可表示为

$$x = \sum_{i=1}^{n} x_i x^{(i)},$$

于是

$$Ax = \sum_{i=1}^{n} x_i Ax^{(i)}.$$

由此,依范数的性质,有

$$\|Ax\|_Y \le \sum_{i=1}^{n} |x_i| \left\|Ax^{(i)}\right\|_Y \le c'\|x\|_\infty, \tag{8.24}$$

其中

$$c' \equiv \sum_{i=1}^{n} \left\|Ax^{(i)}\right\|_Y, \quad \|x\|_\infty = \max_{1\le i\le n}|x_i|.$$

再依范数等价定理 **1.8.13**,存在常数 c'',使得

$$\|x\|_\infty \le c''\|x\|_X,$$

由此及(8.24)推出成立(8.23). □

一般,满足(8.23)的线性算子 A 称为**有界线性算子**.

1.8.20 推论 设 X 和 Y 是两个有限维赋范线性空间,则每个线性算子 $A \in \mathscr{L}(X,Y)$ 是连续的,也就是说,如果

$$\lim_{k\to\infty} x^{(k)} = x,$$

那么

$$\lim_{k\to\infty} Ax^{(k)} = Ax.$$

证 利用(8.23),

$$\left\|Ax^{(k)} - Ax\right\|_Y = \left\|A(x^{(k)} - x)\right\|_Y \le c\left\|x^{(k)} - x\right\|_X, \quad k = 1,2,\cdots$$

即得 A 连续. □

对于线性算子来说,这个推论表明有界性蕴涵连续性,反过来容易证明连续性也蕴涵有界性,因此有界性和连续性两者等价.

1.8.21 定义 设 X 和 Y 是两个有限维赋范线性空间,由

$$\|A\| \equiv \sup_{x\in X, x\neq 0} \frac{\|Ax\|_Y}{\|x\|_X}, \quad \forall A \in \mathscr{L}(X,Y) \tag{8.25}$$

确定 $\mathscr{L}(X,Y)$ 上的一个函数 $\|\bullet\|$,称为**算子范数**.

1.8.19 保证(8.25)中等式右端的上确界存在而且是有限值.

因为范数 $\|\bullet\|_X$ 和 $\|\bullet\|_Y$ 具有齐次性,定义(8.25)可以改写成

$$\|A\| \equiv \sup_{\|x\|_X=1} \|Ax\|_Y, \quad \forall A \in \mathscr{L}(X,Y). \tag{8.26}$$

1.8.22 定理 (8.25)或(8.26)定义的 $\|\bullet\|$ 是 $\mathscr{L}(X,Y)$ 上的范数.

证 设 $A \in \mathscr{L}(X,Y)$ 是非零映射,则存在 $x^{(0)} \in X, x^{(0)} \neq 0$,使得 $Ax^{(0)} \neq 0$.于是由(8.25),

$$\|A\| \geq \frac{\|Ax^{(0)}\|_Y}{\|x^{(0)}\|_X} > 0.$$

表明 $\|\bullet\|$ 具有正定性.

另一方面,$\forall A, B \in \mathscr{L}(X,Y)$,

$$\|A+B\| = \sup_{\|x\|_X=1} \|(A+B)x\|_Y \leq \sup_{\|x\|_X=1} \left(\|Ax\|_Y + \|Bx\|_Y\right)$$

$$\leq \sup_{\|x\|_X=1} \|Ax\|_Y + \sup_{\|x\|_X=1} \|Bx\|_Y = \|A\| + \|B\|.$$

$\|\bullet\|$ 的次加性得证.

$\|\bullet\|$ 满足齐次性是显然的. □

1.8.23 定理 设 X 和 Y 是两个有限维赋范线性空间, X' 和 Y' 是对应的备有对偶范数的对偶空间,且

$$A \in \mathscr{L}(X,Y), \quad A' \in \mathscr{L}(Y',X'),$$

A' 是 A 的转置,则

$$\|A'\| = \|A\|. \tag{8.27}$$

证 依转置定义 **1.3.8**,

$$(A'l, x) = (l, Ax), \quad \forall l \in Y', x \in X.$$

由于 $A'l \in X'$,依对偶范数定义 **1.8.16**,

$$\|A'l\|' = \sup_{\|x\|_X=1}(A'l, x) = \sup_{\|x\|_X=1}(l, Ax) \le \sup_{\|x\|_X=1}\|l\|'\|Ax\|_Y.$$

再依(8.26),

$$\|A'l\|' \le \|l\|' \sup_{\|x\|_X=1}\|Ax\|_Y = \|l\|'\|A\|.$$

由此,得

$$\|A'\| = \sup_{l\neq 0}\frac{\|A'l\|'}{\|l\|'} \le \|A\|. \tag{8.28}$$

在此不等式中用 A' 替换 A ,又有

$$\|A''\| \le \|A'\|. \tag{8.29}$$

注意到 **1.3.8** 中的关系式(3.27), $A'' = A$,而且由 **1.8.18**, X'' 和 Y'' 的范数分别与 X 和 Y 的范数相同.这样,

$$\|A''\| = \|A\|,$$

从而联合(8.28)和(8.29)推出(8.27). □

1.8.24 定理 设 X , Y 和 Z 是三个有限维赋范线性空间,且

$$A \in \mathscr{L}(X,Y), \quad B \in \mathscr{L}(Y,Z),$$

则复合映射 BA 的范数

$$\|BA\| \le \|B\|\|A\|. \tag{8.30}$$

证 依 **1.8.21**,

$$\|By\|_Z \le \|B\|\|y\|_Y, \quad \|Ax\|_Y \le \|A\|\|x\|_X.$$

因此

$$\|BAx\|_Z \le \|B\|\|Ax\|_Y \le \|B\|\|A\|\|x\|_X.$$

从而

$$\|BA\| = \sup_{x \ne 0} \frac{\|BAx\|_Z}{\|x\|_X} \le \|B\|\|A\|. \qquad \square$$

1.8.25 定理 设 X 和 Y 是有限维赋范线性空间, $A \in \mathscr{L}(X,Y)$ 是可逆的, $B \in \mathscr{L}(X,Y)$ 在如下意义下与 A 的差别不是太大:

$$\|B - A\| < c, \quad c \equiv \frac{1}{\|A^{-1}\|}, \tag{8.31}$$

则 B 是可逆的.

证 只须证明线性映射 B 是一对一的和满的.用反证法.假如存在 $x^{(0)} \in X, x^{(0)} \ne 0, Bx^{(0)} = 0$,那么

$$Ax^{(0)} = (A - B)x^{(0)}.$$

因 A 可逆,

$$x^{(0)} = A^{-1}(A - B)x^{(0)}.$$

利用 **1.8.24** 和(8.31),以及 $\|x^{(0)}\|_X > 0$,

$$\|x^{(0)}\|_X \le \|A^{-1}\|\|A - B\|\|x^{(0)}\|_X < \|A^{-1}\|c\|x^{(0)}\|_X = \|x^{(0)}\|_X.$$

得出矛盾.这就证明了 B 是一对一的.

因 A 可逆,它是 X 和 Y 之间的同构映射,故

$$\dim X = \dim Y.$$

这样,从 B 是一对一的及 **1.3.3** 推出 B 是满的. $\square$

1.8.26 定义 设 X 和 Y 是有限维赋范线性空间,线性算子序列

$$\{A_k\} \in \mathscr{L}(X,Y)$$

称为**收敛**于**极限** $A \in \mathscr{L}(X,Y)$,并且记作 $\lim_{k\to\infty} A_k = A$,如果

$$\lim_{k\to\infty}\|A_k - A\| = 0 . \tag{8.32}$$

1.8.27 定理 设 X 是有限维赋范线性空间, $R \in \mathscr{L}(X, X)$ 满足

$$\|R\| < 1 .$$

并设 $B = I - R$,则 B 可逆,而且

$$B^{-1} = \sum_{k=1}^{\infty} R^k . \tag{8.33}$$

这一定理是 **1.8.25** 取 $Y = X, A = I$ 的特殊情形.

(8.33)的证明留作练习.

1.8.28 定义 $m \times n$ 复矩阵全体$\mathbb{C}^{m\times n}$,在矩阵加法和数乘运算下构成复数域$\mathbb{C}$上的线性空间,仍记作$\mathbb{C}^{m\times n}$.

相仿, $m \times n$ 实矩阵全体$\mathbb{R}^{m\times n}$ 构成实数域$\mathbb{R}$上的线性空间. 下面关于$\mathbb{C}^{m\times n}$ 上的说法也可以限制于$\mathbb{R}^{m\times n}$ 上.

设$\|\bullet\|$ 是$\mathbb{C}^{m\times n}$ 上的一个函数,如果成立

(1) 正定性: $\|A\| > 0, \forall A \in \mathbb{C}^{m\times n}, A \neq 0;\quad \|0\| = 0$.

(2) 次加性: $\|A + B\| \le \|A\| + \|B\|, \forall A, B \in \mathbb{C}^{m\times n}$.

(3) 齐次性: $\|\alpha A\| = |\alpha|\,\|A\|, \forall A \in \mathbb{C}^{m\times n}, \alpha \in \mathbb{C}$.

(4) 次乘性: $\|AB\| \le \|A\|\|B\|, \forall A, B \in \mathbb{C}^{m\times n}$.

则称$\|\bullet\|$ 为$\mathbb{C}^{m\times n}$ 上的一个**矩阵范数**.

鉴于 $m \times n$ 矩阵可以视为 mn 维线性空间中的向量,而且对照 **1.8.2**,矩阵范数定义中的前三条性质即为向量范数定义中的三条性质,因此,依 **1.8.5**,以及 **1.8.12** 和 **1.8.13**,可以相仿地建立矩阵序列收敛的概念和矩阵范数等价的概念,并且直接得出下面的结论.

1.8.29 定理 (1) 由

$$\|A\|_F \equiv \left(\sum_{i=1}^{m}\sum_{j=1}^{n}|a_{ij}|^2\right)^{1/2}, \quad \forall A \in \mathbb{C}^{m\times n} \tag{8.34}$$

定义的$\|\bullet\|_F$是$\mathbb{C}^{m\times n}$上的一个矩阵范数,它是复 Euclid 空间中向量的 Euclid 范数(见 **1.7.3**)的推广,称为矩阵的 **Frobenius 范数**或 **Euclid 范数**,也称 **Schur 范数**或 **Hilbert-Schmidt 范数**.

(2) 设

$$\{A^{(k)}\}\subset \mathbb{C}^{m\times n}; \quad A^{(k)} = [a_{ij}^{(k)}], \ k = 1,2,\cdots$$

并设$A = [a_{ij}] \in \mathbb{C}^{m\times n}$,$\|\bullet\|$是$\mathbb{C}^{m\times n}$上任一范数,则成立

$$\lim_{k\to\infty}\|A^{(k)} - A\| = 0 \Leftrightarrow \lim_{k\to\infty} a_{ij}^{(k)} = a_{ij}, \quad i,j = 1,\cdots,n .$$

这就是说,矩阵序列$\{A^{(k)}\}$依任一范数收敛于A等价于$\{A^{(k)}\}$的每个对应元素的序列$\{a_{ij}^{(k)}\}$收敛于A的对应元素a_{ij}.

(3) $\mathbb{C}^{m\times n}$上的所有范数是等价的.

证明留作练习.

1.8.30 定义 设$\|\bullet\|_{\mathbb{C}^n}$是$\mathbb{C}^n$上的一个范数.由

$$\|A\| \equiv \sup_{x\in\mathbb{C}^n, x\neq 0} \frac{\|Ax\|_{\mathbb{C}^n}}{\|x\|_{\mathbb{C}^n}}, \quad \forall A \in \mathbb{C}^{n\times n}, \tag{8.35}$$

确定$\mathbb{C}^{n\times n}$上的一个函数$\|\bullet\|$,称为矩阵的由向量范数$\|\bullet\|_{\mathbb{C}^n}$导出的**算子范数**.

由于$n\times n$矩阵与$\mathscr{L}(\mathbb{C}^n,\mathbb{C}^n)$的线性算子一一对应,是同一的,因此矩阵算子范数概念蕴涵在一般线性映射算子范数的定义 **1.8.21** 之中,线性映射有关算子范数的结论都可应用于矩阵算子范数.

特别,**1.8.19** 保证(8.35)是有意义的;**1.8.22** 和 **1.8.24** 表明矩阵算子范数满足矩阵范数定义 **1.8.28** 中的四条性质.

1.8.31 定理 由 **1.8.8** 定义的向量∞－范数,1－范数和2－范数导出的矩阵算子范数如下:

$\forall A=[a_{ij}]\in\mathbb{C}^{n\times n}$,

∞－**算子范数**也称**行范数**:

$$\|A\|_\infty=\sup_{x\neq 0}\frac{\|Ax\|_\infty}{\|x\|_\infty}=\max_{1\le i<n}\sum_{j=1}^{n}|a_{ij}|. \tag{8.36}$$

1－**算子范数**也称**列范数**:

$$\|A\|_1=\sup_{x\neq 0}\frac{\|Ax\|_1}{\|x\|_1}=\max_{1<j\le n}\sum_{i=1}^{n}|a_{ij}|. \tag{8.37}$$

2－**算子范数**常称**谱范数**:

$$\|A\|_2=\sup_{x\neq 0}\frac{\|Ax\|_2}{\|x\|_2}=\sigma_1. \tag{8.38}$$

其中 σ_1 是 A 的最大奇异值,即 A^*A 的最大特征值的非负平方根.

此定理的证明在许多数值分析教科书中可以找到.(8.36)和(8.37)的证明从略,可作练习.(8.38)的证明见 **3.6.8**.

1.8.32 定义 $\mathbb{C}^n$ 上的一个向量范数 $\|\bullet\|_{\mathbb{C}^n}$ 称为和 $\mathbb{C}^{n\times n}$ 上的一个矩阵范数 $\|\bullet\|$ **相容**,有时也称 $\|\bullet\|_{\mathbb{C}^n}$ **从属** $\|\bullet\|$,如果成立

$$\|Ax\|_{\mathbb{C}^n}\le\|A\|\|x\|_{\mathbb{C}^n},\quad \forall x\in\mathbb{C}^n,\ A\in\mathbb{C}^{n\times n}. \tag{8.39}$$

从 **1.8.30** 得知,向量范数 $\|\bullet\|_{\mathbb{C}^n}$ 和由其导出的矩阵算子范数 $\|\bullet\|$ 是相容的.这表明每个向量范数必存在与之相容的矩阵范数.下一定理是说其逆命题亦成立,即每个矩阵范数必存在与之相容的向量范数.

1.8.33 定理 设$\|\cdot\|$是$\mathbb{C}^{n\times n}$上的一个矩阵范数,则$\mathbb{C}^n$上存在和$\|\cdot\|$相容的向量范数.

证 定义

$$\|x\|_{\mathbb{C}^n} \equiv \|[x \quad 0 \quad \cdots \quad 0]\|, \quad \forall x \in \mathbb{C}^n,$$

其中$[x \quad 0 \quad \cdots \quad 0] \in \mathbb{C}^{n\times n}$是第1列为$x$其余列为零向量的矩阵.容易验证$\|\cdot\|_{\mathbb{C}^n}$是$\mathbb{C}^n$上的向量范数,而且满足

$$\begin{aligned}\|Ax\|_{\mathbb{C}^n} &= \|[Ax \quad 0 \quad \cdots \quad 0]\| = \|A[x \quad 0 \quad \cdots \quad 0]\| \\ &\le \|A\|\|[x \quad 0 \quad \cdots \quad 0]\| = \|A\|\|x\|_{\mathbb{C}^n}.\end{aligned}$$

□

1.9 凸性的基本概念

1.9.1 定义 设X是$\mathbb{R}$上的线性空间,$x, y \in X$,点集

$$\{\alpha x + (1-\alpha)y : 0 \le \alpha \le 1\} \tag{9.1}$$

称为X的以x和y为端点的**线段**.

显然,线性空间包含所有以其元素为端点的线段.

注意:本节以下均假设X是$\mathbb{R}$上的线性空间.

1.9.2 定义 设集合$K \subset X$.如果对于任何$x, y \in K$,以x和y为端点的线段上的所有点也都属于K,亦即

$$\{\alpha x + (1-\alpha)y : 0 \le \alpha \le 1\} \subset K, \quad \forall x, y \in K, \tag{9.2}$$

则称K为X中的**凸集**.

直观上,K是凸集,应该没有"凹部","裂痕"和"洞孔".

凸性是一个很本原的概念,所基于的除实数域上线性空间结构

的简单明了的要素而外别无其它.然而它的某些基本结果却异常深刻,并因此使这些结果出现在十分广泛的论题之中.

例 (1) 如下集合均为凸集:整个空间 X 、空集 O 、单点集 $\{x\}$ 、X 上的任一线段.

(2) $\mathbb{R}$ 上所有实系数多项式构成的线性空间中,在区间 $(0,1)$ 上为正的所有多项式构成的子集是凸集.

(3) 实自伴矩阵(见第 3 章)全体构成的线性空间中,正矩阵(见第 4 章)全体构成的子集是凸集.

以上各例的证明留作练习.

1.9.3 定义 设 l 是 X 上的线性函数, $c \in \mathbb{R}$.

(1) 点集 $\{x : l(x) = c, \forall x \in X\}$ 称为 X 的**超平面**.

(2) 点集 $\{x : l(x) < c, \forall x \in X\}$ 称为 X 的**开半空间**.

(3) 点集 $\{x : l(x) \leq c, \forall x \in X\}$ 称为 X 的**闭半空间**.

显然,超平面、开和闭半空间均是凸集.

1.9.4 定理 (1) 任意多个凸集的交集是凸集.

(2) 设 $K, H \subset X$ 是凸集,则 K 与 H 的**和集**

$$K + H \equiv \{x + y : x \in K, y \in H\}$$

是凸集.

这两个结论可从 **1.9.2** 直接推出,留作练习.

利用这一定理,从少数基本凸集出发可以构造出种种凸集.

1.9.5 定义 设集合 $S \subset X$,点 $x \in S$.如果对于每个点 $y \in X$,存在常数 $\varepsilon_y > 0$,使得

$$x + ty \in S, \quad \forall t \in (0, \varepsilon_y), \tag{9.3}$$

则称 x 为 S 的**内点**.

1.9.6 定义 如果凸集 K 的每一点都是其内点,则称 K 为**开凸集**.

显然,开半空间是开凸集.

1.9.7 定义 如果凸集 K 对任何 $x \in K$ 成立

$$\alpha x \in K, \quad \forall \alpha > 0, \tag{9.4}$$

则称 K 为**凸锥**.

容易证明(留作练习):两个凸锥的交集与和集仍然是凸锥.

1.9.8 定义 设 $K \subset X$ 是包含零元素的开凸集.由

$$p_k(x) \equiv \inf\left\{r : r > 0, \frac{x}{r} \in K\right\}, \quad \forall x \in X \tag{9.5}$$

确定的函数 $p_k : X \to \mathbb{R}$ 称为 K 的**规范函数**(gauge function).

1.9.9 定理 设 p_k 是包含零元素的开凸集 $K \subset X$ 的规范函数,则

(1) p_k 对每点 $x \in X$ 是有定义的.

(2) p_k 是正齐次的:

$$p_k(\alpha x) = \alpha p_k(x), \quad \forall \alpha > 0. \tag{9.6}$$

(3) p_k 是次加的:

$$p_k(x + y) \le p_k(x) + p_k(y), \quad \forall x, y \in X. \tag{9.7}$$

(4) $p_k(x) < 1$ 的充分必要条件是 $x \in K$.

证 (1) 由于 $0 \in K$, 0 是 K 的内点,依 **1.9.5**,对每个 $x \in X$,当 $r > 0$ 充分大时,有 $\frac{1}{r} x \in K$,因此集合

$$\left\{r : r > 0, \frac{x}{r} \in K\right\}$$

非空且以零为下界,故其必有下确界.

(2) 若 $\frac{x}{r} \in K$,则 $\frac{\alpha x}{\alpha r} \in K$.由此及(9.5)即推出(9.6).

(3) 设 s 和 t 是正实数,使得

$$p_k(x) < s, \quad p_k(y) < t, \tag{9.8}$$

则依(9.5)及 $0 \in K$,推出 $\frac{x}{s}, \frac{y}{t} \in K$.这样,因 K 是凸的,$\frac{x}{s}$ 和 $\frac{y}{t}$ 连线上的点

$$\frac{x+y}{s+t} \equiv \frac{s}{s+t}\frac{x}{s} + \frac{t}{s+t}\frac{y}{t} \tag{9.9}$$

必属于 K.于是,仍依(9.5),

$$p_k(x+y) \le s+t. \tag{9.10}$$

由此并因 s 和 t 的选取可以分别地任意趋近 $p_k(x)$ 和 $p_k(y)$,推出成立(9.7).

(4) 假设 $p_k(x) < 1$,依(9.5),存在 $r < 1$,使得 $\frac{x}{r} \in K$.恒等式

$$x \equiv r\frac{x}{r} + (1-r)0,$$

表明 x 在以 0 和 $\frac{x}{r}$ 为端点的线段上,鉴于 K 是凸的,必有 $x \in K$.

反之,假设 $x \in K$,则 x 是 K 的内点.因此,选取足够小的 $\varepsilon > 0$,有 $x + \varepsilon x \in K$.

令 $r = 1/(1+\varepsilon)$,于是

$$\frac{x}{r} = (1+\varepsilon)x \in K.$$

依(9.5),

$$p_k(x) \le r = \frac{1}{1+\varepsilon}. \qquad \square$$

这一定理给出了开凸集的解析摹状.还有另一种对偶摹状,可参见[L1997].

1.9.10 定义 凸集 $K \subset X$ 称为**闭的**,如果属于 K 的每一开线段

$$\{\alpha x + (1-\alpha)y : 0 < \alpha < 1\} \tag{9.11}$$

都有其端点 $x, y \in K$.

容易证明(留作练习):闭凸集的交集仍是闭凸集.

例 如下集合均为闭凸集: 整个空间 X、空集 O、单点集 $\{x\}$、X 上的任一闭线段 $\{\alpha x + (1-\alpha)y : 0 \le \alpha \le 1\}$.

1.9.11 定义 设集合 $S \subset X$,包含 S 的所有闭凸集的交集称为 S 的**闭凸包**.

1.9.12 定义 设 $x^{(1)}, \cdots, x^{(m)} \in X$, $\alpha_1, \cdots, \alpha_m \in \mathbb{R}$ 且满足

$$\alpha_i \ge 0, \quad i = 1, \cdots, m; \quad \alpha_1 + \cdots + \alpha_m = 1, \tag{9.12}$$

则称

$$x \equiv \sum_{i=1}^{m} \alpha_i x^{(i)} \tag{9.13}$$

为 $x^{(1)}, \cdots, x^{(m)}$ 的**凸组合**.

不难证明(留作练习):若 $x^{(1)}, \cdots, x^{(m)}$ 属于凸集 K,则它们的任何凸组合均属于 K.

1.9.13 定义 设 $K \subset X$ 是凸集,点 $x \in X$ 称为 K 的**边界点**,如果 $x \in K$,但 x 不是 K 的内点.

1.9.14 定义 设$K \subset X$是凸集，点$x \in K$称为K的**极值点**,如果x不是K中线段的内点;换句话说,

$$x = \alpha y + (1-\alpha)z, \quad 0 < \alpha < 1; \quad y, z \in K,$$

蕴涵$y = z = x$.

由定义知,K的所有极值点是K的边界点,但边界点并非必是极值点.例如,K是闭的凸多边形,所有边上的点和顶点是边界点,但只有顶点是极值点.

1.9.15 定义 凸集$K \subset X$称为有界的,如果它不包含形如

$$\{x + ty : t \geq 0\}, \quad x, y \in X, \quad x \neq y \tag{9.14}$$

的射线.

1.9.16 Carathéodory 定理 设$\dim X = n$,并设$K \subset X$是有界闭凸集，则K的每一点可以表示作K的至多$n+1$个极值点的凸组合.

证明从略,可参见[L1997].

1.9.17 定义 设$K \subset X$是凸集，映射

$$f : K \to \mathbb{R}$$

称为(K上的)**凸函数**,如果成立

$$f(\alpha x + (1-\alpha)y) \leq \alpha f(x) + (1-\alpha)f(y), \quad 0 < \alpha < 1;$$
$$\forall x, y \in K, \quad x \neq y. \tag{9.15}$$

f称为**严格凸函数**,如果(9.15)的不等式成立严格不等号.

f称为**凹函数**,如果$-f$是凸函数.

f称为**严格凹函数**,如果$-f$是严格凸函数.

几何上,f是K上的凸(凹)函数,对于任意的$x, y \in K$,连接点$(x, f(x))$与点$(y, f(y))$的弦,总在f在线段

$$\{\alpha x+(1-\alpha)y:0<\alpha<1\}$$

上的图形之上(下).

1.9.18 述评 对于 $X=\mathbb{R}^n$ 和 $K\subset\mathbb{R}^n$ 是开凸集的情形,凸函数 $f:K\to\mathbb{R}$的 **Hesse 矩阵**

$$H(x)\equiv\left[\frac{\partial^2 f}{\partial x_i\partial x_j}(x)\right], \tag{9.16}$$

在 K 上是几乎处处有定义的;在有定义的点 $x\in K$ 处,$H(x)$是对称的,而且必是半正定的.当 f 是严格凸时 $H(x)$是正定的.反之,如果 $H(x)$在 K 上半正定(正定),那么 f 是凸的(严格凸的).类似地,负定性对应于凹性 (正定性和负定性见第 3 章).

凸和凹函数的最优化有一些引入的性质.在紧凸集上,凸(凹)函数在极值点达到最小值(最大值).另一方面,在凸集上,使凸(凹)函数达到最小值(最大值)的点集是凸的,而且任何局部最小值(最大值)是全局最小值(最大值).例如,严格凸函数至多在凸集的一个点处达到最小值,而且临界点必定是取最小值的点.

实数的凸组合遵从一些简单但经常有用的不等式.

若 $x_1,\cdots,x_n\in\mathbb{R}$,则

$$\min_{1\le i\le n}x_i\le\sum_{i=1}^n\alpha_i x_i\le\max_{1\le i\le n}x_i,$$

$$\forall\alpha_1,\cdots,\alpha_n\ge 0,\quad \alpha_1+\cdots+\alpha_n=1. \tag{9.17}$$

考虑某些简单的定义在区间 I 上的一元凸函数 f,用以导出各种经典不等式.用数学归纳法可以证明:在 $K=I$ 上的两点不等式(9.15)等价于 n 点不等式

$$f\left(\sum_{i=1}^n\alpha_i x_i\right)\le\sum_{i=1}^n\alpha_i f(x_i);$$

$$\forall \alpha_1,\cdots,\alpha_n \ge 0,\quad \alpha_1+\cdots+\alpha_n = 1,$$

$$\forall x_1,\cdots,x_n \in I;\quad n = 2,3,\cdots \tag{9.18}$$

应用(9.18)于$I=(0,+\infty)$上的严格凸函数$f(x)=-\log x$,导出**带权算术-几何平均不等式**

$$\sum_{i=1}^{n}\alpha_i x_i \ge \prod_{i=1}^{n} x_i^{\alpha_i};\ \forall x_1,\cdots,x_n \ge 0,$$

$$\forall \alpha_1,\cdots,\alpha_n \ge 0,\quad \alpha_1+\cdots+\alpha_n = 1. \tag{9.19}$$

特别,当$\alpha_1=\cdots=\alpha_n=\dfrac{1}{n}$时,成为通常**算术-几何平均不等式**

$$\frac{1}{n}\sum_{i=1}^{n} x_i \ge \left(\prod_{i=1}^{n} x_i\right)^{1/n},\ \forall x_1,\cdots,x_n \ge 0, \tag{9.20}$$

等号当且仅当$x_1=\cdots=x_n$时成立.

应用(9.18)于$I=(0,+\infty)$上的幂函数$f(x)=x^p, p>1$,导出**Hölder 不等式**

$$\sum_{i=1}^{n} x_i y_i \le \left(\sum_{i=1}^{n} x_i^p\right)^{1/p}\left(\sum_{i=1}^{n} y_i^q\right)^{1/q};\quad p>1,\quad \frac{1}{p}+\frac{1}{q}=1,$$

$$\forall x_1,\cdots,x_n \ge 0,\ y_1,\cdots,y_n \ge 0, \tag{9.21}$$

等号当且仅当$(x_1^p,\cdots,x_n^p)$和$(y_1^q,\cdots,y_n^q)$线性相关时成立.特别,当$p=q=2$时,成为 **Cauchy-Schwarz 不等式**

$$\sum_{i=1}^{n} x_i y_i \le \left(\sum_{i=1}^{n} x_i^2\right)^{1/2}\left(\sum_{i=1}^{n} y_i^2\right)^{1/2};$$

$$\forall x_1,\cdots,x_n \ge 0,\ y_1,\cdots,y_n \ge 0, \tag{9.22}$$

等号当且仅当$(x_1,\cdots,x_n)$和$(y_1,\cdots,y_n)$线性相关时成立.

从(9.21)可以推出 **Minkowski 不等式**

$$\left(\sum_{i=1}^{n}(x_i+y_i)^p\right)^{1/p} \le \left(\sum_{i=1}^{n}x_i^p\right)^{1/p} + \left(\sum_{i=1}^{n}y_i^p\right)^{1/p}, \quad p>1,$$

$$\forall x_1,\cdots,x_n \ge 0,\ y_1,\cdots,y_n \ge 0, \qquad (9.23)$$

等号当且仅当$(x_1,\cdots,x_n)$和$(y_1,\cdots,y_n)$线性相关时成立.

2 基本性质矩阵

2.1 若干基本术语和矩阵

2.1.1 定义 设 $A=[a_{ij}]\in\mathbb{C}^{m\times n}$,**指标集**

$$\alpha\equiv\{i_1,\cdots,i_k\}\subset\{1,\cdots,m\},\quad \beta\equiv\{j_1,\cdots,j_l\}\subset\{1,\cdots,n\}. \tag{1.1}$$

记号 $A(\alpha,\beta)$表示 A 的**子矩阵**,

$$A(\alpha,\beta)\equiv\begin{bmatrix} a_{i_1j_1} & a_{i_1j_2} & \cdots & a_{i_1j_l} \\ a_{i_2j_1} & a_{i_2j_2} & \cdots & a_{i_2j_l} \\ \cdots & \cdots & \cdots & \cdots \\ a_{i_kj_1} & a_{i_kj_2} & \cdots & a_{i_kj_l} \end{bmatrix}\in\mathbb{C}^{k\times l}. \tag{1.2}$$

若 $A\in\mathbb{C}^{n\times n}$,$\alpha\subset\{1,\cdots,n\}$,则称子矩阵 $A(\alpha,\alpha)$为 A 的**主子矩阵**(principal submatrix),并简记作 $A(\alpha)$.特别

$$A(\{1,\cdots,i\})=\begin{bmatrix} a_{11} & \cdots & a_{1i} \\ \cdots & \cdots & \cdots \\ a_{i1} & \cdots & a_{ii} \end{bmatrix},\quad 1\le i\le n, \tag{1.3}$$

称为 A 的**前主子矩阵**(leading principal submatrix).

当子矩阵(1.2)为方阵,即 $k=l$ 时,其行列式

$$\det A(\alpha,\beta) \tag{1.4}$$

称为 A 的**子式**.主子矩阵 $A(\alpha)$的行列式 $\det A(\alpha)$称为 A 的**主子式**.特别,前主子矩阵的行列式

$$\det A(\{1,\cdots,i\}),\quad 1\le i\le n \tag{1.5}$$

称为 A 的**前主子式**.

例 设

$$A=\begin{bmatrix}9&3&0\\7&5&4\\8&0&5\\6&2&1\\0&4&2\end{bmatrix},\quad B=\begin{bmatrix}1&2&3\\4&5&6\\7&8&9\end{bmatrix} \tag{1.6}$$

A 的子矩阵和子式举例如下:

$$A(\{1,2,4\},\{1,3\})=\begin{bmatrix}9&0\\7&4\\6&1\end{bmatrix},$$

$$\det A(\{1,4\},\{2,3\})=\begin{vmatrix}3&0\\2&1\end{vmatrix}=3.$$

B 的前主子矩阵为

$$B(\{1\})=[1],\quad B(\{1,2\})=\begin{bmatrix}1&2\\4&5\end{bmatrix},\quad B(\{1,2,3\})=B.$$

2.1.2 定义 设 $A\in\mathbb{C}^{m\times n}$.

所谓 $\alpha_1,\cdots,\alpha_t$ 构成 $\{1,\cdots,m\}$ 的一种**划分**,是指 $\alpha_1,\cdots,\alpha_t$ 为将集合 $\{1,\cdots,m\}$ 保持 $1,\cdots,m$ 顺序进行某种划分而得的 t 个子集.

现设 $\alpha_1,\cdots,\alpha_t$ 构成 $\{1,\cdots,m\}$ 的一种划分,而且 $\beta_1,\cdots,\beta_s$ 构成 $\{1,\cdots,n\}$ 的一种划分.则称

$$\{A(\alpha_i,\beta_j):1\le i\le t,1\le j\le s\} \tag{1.7}$$

为 A 的一种**分块**.相应地,A 有**分块表示**形式

$$A=\left[A(\alpha_i,\beta_j)\right]$$

$$\equiv\begin{bmatrix} A(\alpha_1,\beta_1) & A(\alpha_1,\beta_2) & \cdots & A(\alpha_1,\beta_s) \\ A(\alpha_2,\beta_1) & A(\alpha_2,\beta_2) & \cdots & A(\alpha_2,\beta_s) \\ \cdots & \cdots & \cdots & \cdots \\ A(\alpha_t,\beta_1) & A(\alpha_t,\beta_2) & \cdots & A(\alpha_t\beta_s) \end{bmatrix}. \tag{1.8}$$

设 $A,B\in\mathbb{C}^{m\times n}$.显然,如果 A 和 B 的分块是**叠合**的,亦即

$$A=\left[A(\alpha_i,\beta_j)\right],\quad B=\left[B(\alpha_i,\beta_j)\right],$$

则成立**分块加法**

$$A+B=\left[A(\alpha_i,\beta_j)+B(\alpha_i,\beta_j)\right]. \tag{1.9}$$

2.1.3 定义 $A\in\mathbb{C}^{m\times n}$ 和 $B\in\mathbb{C}^{n\times p}$ 的分块表示称为**可相乘**的,如果 A 和 B 关于 $\{1,\cdots,n\}$ 的划分是叠合的.

如果 $A=\left[A(\alpha_i,\beta_j)\right]$ 和 $B=\left[B(\alpha_i,\beta_j)\right]$ 是可相乘的分块表示,而且

$$\{1,\cdots,m\}\ \text{划分为}\ \alpha_1,\cdots,\alpha_t,$$
$$\{1,\cdots,n\}\ \text{划分为}\ \beta_1,\cdots,\beta_s,$$
$$\{1,\cdots,p\}\ \text{划分为}\ \gamma_1,\cdots,\gamma_r,$$

则成立**分块乘法**

$$AB=\left[(AB)(\alpha_i,\gamma_j)\right], \tag{1.10}$$

其中

$$(AB)(\alpha_i,\gamma_j)=\sum_{k=1}^{n}A(\alpha_i,\beta_k)B(\beta_k,\gamma_j),$$

$$i=1,\cdots,t,\quad j=1,\cdots,r. \tag{1.11}$$

证明留作练习.

例 A 和 B 如(1.6),分块表示为

$$A=\begin{bmatrix} A(\{1,2\},\{1,2\}) & A(\{1,2\},\{3\}) \\ A(\{3,4,5\},\{1,2\}) & A(\{3,4,5\},\{3\}) \end{bmatrix} \equiv \begin{bmatrix} A_1 & A_2 \\ A_3 & A_4 \end{bmatrix},$$

和

$$B=\begin{bmatrix} B(\{1,2\},\{1,2,3\}) \\ B(\{3\},\{1,2,3\}) \end{bmatrix} \equiv \begin{bmatrix} B_1 \\ B_2 \end{bmatrix},$$

其中

$$A_1=\begin{bmatrix} 9 & 3 \\ 7 & 5 \end{bmatrix},\quad A_2=\begin{bmatrix} 0 \\ 4 \end{bmatrix},\quad A_3=\begin{bmatrix} 8 & 0 \\ 6 & 2 \\ 0 & 4 \end{bmatrix},\quad A_4=\begin{bmatrix} 5 \\ 1 \\ 2 \end{bmatrix},$$

$$B_1=\begin{bmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \end{bmatrix},\quad B_2=\begin{bmatrix} 7 & 8 & 9 \end{bmatrix}.$$

A 和 B 的分块表示是可相乘的,于是有

$$AB=\begin{bmatrix} A_1B_1+A_2B_2 \\ A_3B_1+A_4B_2 \end{bmatrix}$$

$$=\begin{bmatrix} \begin{bmatrix} 21 & 33 & 45 \\ 27 & 39 & 51 \end{bmatrix}+\begin{bmatrix} 0 & 0 & 0 \\ 28 & 32 & 36 \end{bmatrix} \\ \begin{bmatrix} 8 & 16 & 24 \\ 14 & 22 & 30 \\ 16 & 20 & 24 \end{bmatrix}+\begin{bmatrix} 35 & 40 & 45 \\ 7 & 8 & 9 \\ 14 & 16 & 18 \end{bmatrix} \end{bmatrix}=\begin{bmatrix} 21 & 33 & 45 \\ 55 & 71 & 87 \\ 43 & 56 & 69 \\ 21 & 30 & 39 \\ 30 & 36 & 42 \end{bmatrix}.$$

2.1.4 定义 设 $A\in\mathbb{C}^{m\times n}$ 形如

$$A=\begin{bmatrix} A_{11} & 0 & \cdots & 0 \\ 0 & \ddots & \ddots & \vdots \\ \vdots & \ddots & \ddots & 0 \\ 0 & \cdots & 0 & A_{ss} \end{bmatrix} \equiv \operatorname{diag}(\mathrm{A}_{11},\cdots,A_{ss}), \quad (1.12)$$

其中

$$A_{ii}\in\mathbb{C}^{n_i\times n_i},\ i=1,\cdots,s;\qquad n_1+\cdots+n_s=n,$$

则称 A 为**块对角矩阵**.如此矩阵也常记作

$$A=A_{11}\oplus\cdots\oplus A_{ss},\tag{1.13}$$

并称 A 为 $A_{11},\cdots,A_{ss}$ 的**直和**.

关于矩阵的直和这一术语,源自线性空间上的线性算子可以分裂成若干不变子空间上线性算子的直和.可参看[LT1985].

2.1.5 定义 设 $\{a^{(1)},\cdots,a^{(n)}\}$ 是 $\mathbb{C}^n$ 中的一组基,并设向量 $u\in\mathbb{C}^n$ 成立线性表出

$$u=\sum_{i=1}^{n}x_i a^{(i)},\quad x_1,\cdots,x_n\in\mathbb{C},\tag{1.14}$$

则称 $x\equiv(x_1,\cdots,x_n)^{\mathrm{T}}$ 为向量 u **关于基** $\{a^{(1)},\cdots,a^{(n)}\}$ **的表示**,并且称 $x_1,\cdots,x_n$ 为 u 关于该基的**坐标**或**分量**.

通常,向量 $x=(x_1,\cdots,x_n)^{\mathrm{T}}\in\mathbb{C}^n$,指 $(x_1,\cdots,x_n)^{\mathrm{T}}$ 是 x 关于标准基 $e_1,\cdots,e_n$ 的表示.

容易证明(留作练习): $\mathbb{C}^n$ 中每个向量关于取定的基的表示是唯一的.

2.1.6 定义 设 $\{a^{(1)},\cdots,a^{(n)}\}$ 和 $\{b^{(1)},\cdots,b^{(n)}\}$ 是 $\mathbb{C}^n$ 中的两组基.如果 $a^{(j)}$ 关于基 $\{b^{(1)},\cdots,b^{(n)}\}$ 的表示为 $(c_{1j},\cdots,c_{nj})^{\mathrm{T}}$,即成立

$$a^{(j)}=\sum_{i=1}^{n}c_{ij}b^{(i)},\quad j=1,\cdots,n.\tag{1.15}$$

则称矩阵 $C\equiv[c_{ij}]\in\mathbb{C}^{n\times n}$ 为从基 $\{a^{(1)},\cdots,a^{(n)}\}$ 到基 $\{b^{(1)},\cdots,b^{(n)}\}$ 的**转移矩阵**(transition matrix).

容易证明(留作练习):转移矩阵是非奇异的而且是唯一的.

2.1.7 定理 设$\mathbb{C}^n$中从基$\{a^{(1)},\cdots,a^{(n)}\}$到基$\{b^{(1)},\cdots,b^{(n)}\}$的转移矩阵为$C\equiv[c_{ij}]\in\mathbb{C}^{n\times n}$,向量$u\in\mathbb{C}^n$关于两组基的表示依次是$x\equiv(x_1,\cdots,x_n)^{\mathrm{T}}$和$y\equiv(y_1,\cdots,y_n)^{\mathrm{T}}$.则

$$y=Cx. \tag{1.16}$$

证 依假设成立(1.15),而且

$$u=\sum_{j=1}^{n}x_j a^{(j)}=\sum_{i=1}^{n}y_i b^{(i)}$$

利用(1.15),

$$\sum_{j=1}^{n}x_j a^{(j)}=\sum_{j=1}^{n}x_j\left(\sum_{i=1}^{n}c_{ij}b^{(i)}\right)=\sum_{i=1}^{n}\left(\sum_{j=1}^{n}c_{ij}x_j\right)b^{(i)}$$

所以有

$$u=\sum_{i=1}^{n}y_i b^{(i)}=\sum_{i=1}^{n}\left(\sum_{j=1}^{n}c_{ij}x_j\right)b^{(i)}$$

于是,由关于基的表示的唯一性,推出

$$y_i=\sum_{j=1}^{n}c_{ij}x_j,\quad i=1,\cdots,n$$

此即(1.16). □

2.1.8 定义 设$A\in\mathbb{C}^{m\times n}$.$A$的同一阶的所有子式的阵列称为$A$的**合成矩阵**(compound matrix).

特别,(α,β)-元素是$\det A(\alpha,\beta)$的$\binom{m}{k}\times\binom{n}{k}$阵列称为$A$的**第$k$合成矩阵**,记作$C_k(A)$.这里

$$\alpha\subset\{1,\cdots,m\},\quad \beta\subset\{1,\cdots,n\}$$

是**基数**$k\le\min\{m,n\}$的指标集,并按通常字典排序.比如$k=3$,指

标集$\{1,2,4\}$在$\{1,2,5\}$之前,$\{1,2,5\}$又在$\{1,3,4\}$之前,等等.

例 设

$$A=\begin{bmatrix}1&0&2\\0&3&0\\4&0&5\end{bmatrix},$$

则

$$C_2(A)=\begin{bmatrix}\det\begin{bmatrix}1&0\\0&3\end{bmatrix}&\det\begin{bmatrix}1&2\\0&0\end{bmatrix}&\det\begin{bmatrix}0&2\\3&0\end{bmatrix}\\\det\begin{bmatrix}1&0\\4&0\end{bmatrix}&\det\begin{bmatrix}1&2\\4&5\end{bmatrix}&\det\begin{bmatrix}0&2\\0&5\end{bmatrix}\\\det\begin{bmatrix}0&3\\4&0\end{bmatrix}&\det\begin{bmatrix}0&0\\4&5\end{bmatrix}&\det\begin{bmatrix}3&0\\0&5\end{bmatrix}\end{bmatrix}$$

$$=\begin{bmatrix}3&0&-6\\0&-3&0\\-12&0&15\end{bmatrix}.$$

2.1.9 定理 合成矩阵有如下性质:

(1) 若$A\in\mathbb{C}^{m\times l}$,$B\in\mathbb{C}^{l\times n}$,则

$$C_k(AB)=C_k(A)C_k(B),\quad k\le\min\{l,m,n\}.$$

(2) $C_k(\alpha A)=\alpha^kC_k(A),\alpha\in\mathbb{C}$.

(3) 若单位矩阵$I\in\mathbb{C}^{n\times n}$,则

$$C_k(I)=I\in\mathbb{C}^{\binom{n}{k}\times\binom{n}{k}}.$$

(4) 若$A\in\mathbb{C}^{n\times n}$是非奇异的,则

$$C_k(A)^{-1}=C_k(A^{-1}).$$

(5) 若$A\in\mathbb{C}^{m\times n}$,则

$$C_k(A)^{\mathrm{T}} = C_k\left(A^{\mathrm{T}}\right).$$

(6) 若 $A \in \mathbb{C}^{m\times n}$,则

$$C_k(A)^* = C_k\left(A^*\right).$$

证明留作练习.

2.1.10 定义 对角矩阵 $D \in \mathbb{C}^{n\times n}$ 称为**纯量矩阵**(scalar matrix),如果 D 的对角元素全是相等的,也就是说,D 形如

$$D = \alpha I, \quad \alpha \in \mathbb{C}.$$

显然,用纯量矩阵 $D = \alpha I \in \mathbb{C}^{n\times n}$ 左乘或右乘任何矩阵 $A \in \mathbb{C}^{n\times n}$,其效果均等同于用数 α 乘 A,即 $DA = AD = \alpha A$.因此,纯量矩阵和任何同阶矩阵(相乘)是可交换的.

2.1.11 定义 矩阵

$$E \equiv \begin{bmatrix} e_n & \cdots & e_1 \end{bmatrix} = \begin{bmatrix} 0 & \cdots & 0 & 1 \\ \vdots & \iddots & \iddots & 0 \\ 0 & \iddots & \iddots & \vdots \\ 1 & 0 & \cdots & 0 \end{bmatrix}, \tag{1.17}$$

称为 n 阶**反序单位矩阵**(reverse unit matrix),或**后向单位矩阵**(backward identity matrix),或**交换矩阵**(exchange matrix),或**翻转矩阵**(flip matrix).

显然,

$$E^2 = I. \tag{1.18}$$

2.1.12 定义 设 $A = [a_{ij}]$ 是 $m \times n$ 矩阵.

如果 $a_{ij} = 0, \forall i > j$,则称 A 为**上三角矩阵**.

如果 $a_{ij} = 0, \forall i \geq j$,则称 A 为**严格上三角矩阵**.

如果$a_{ij}=0, \forall j>i$,则称A为**下三角矩阵**.

如果$a_{ij}=0, \forall j \geq i$,则称$A$为**严格下三角矩阵**.

如果A是(上或下)三角矩阵,而且其对角元素全为1,则称A为**单位(上或下)三角矩阵**.

容易证明(留作练习):

(1) 两个可相乘的上(或下)三角矩阵之积仍是上(或下)三角矩阵.

(2) $n \times n$三角矩阵$A=[a_{ij}]$的全部特征值是其对角元素$a_{11},\cdots,a_{nn}$.

2.1.13 定义 $A \in \mathbb{C}^{n\times n}$称为**本性三角矩阵**(essentially triangular matrix),如果存在置换矩阵P,使得PAP^{T}是三角矩阵.

2.1.14 定义 设$A \in \mathbb{C}^{n\times n}$形如

$$A=\begin{bmatrix} A_{11} & A_{12} & \cdots & A_{1s} \\ 0 & A_{22} & \cdots & A_{2s} \\ \vdots & \ddots & \ddots & \vdots \\ 0 & \cdots & 0 & A_{ss} \end{bmatrix}, \tag{1.19}$$

其中

$$A_{ii} \in \mathbb{C}^{n_i \times n_i}, i=1,\cdots,s, n_1+\cdots+n_s=n,$$

则称A为**块上三角矩阵**.

类似地,可以定义**块下三角矩阵**,**严格块上三角矩阵**,**严格块下三角矩阵**.

容易证明(留作练习):若A形如(1.19),则成立

$$\det A=\prod_{i=1}^{s}\det A_{ii} \tag{1.20}$$

和

$$\operatorname{rank}A \geq \sum_{i=1}^{s} \operatorname{rank}A_{ii}. \tag{1.21}$$

2.1.15 定义 设 $A \in \mathbb{C}^{n\times n}$,如果成立

$$A^2 = A, \tag{1.22}$$

则称 A 为**等幂矩阵**(idempotent matrix).

容易证明(留作练习):对角矩阵为等幂矩阵的充分必要条件是其每个对角元素等于0或1.

2.1.16 定义 设 $A \in \mathbb{C}^{n\times n}$,如果存在正整数 k,使得

$$A^k = 0, \tag{1.23}$$

则称 A 为 k **次幂零矩阵**(nilpotent matrix).

容易证明(留作练习):若 $A \in \mathbb{C}^{n\times n}$ 是严格(上或下)三角矩阵,则 A 是 n 次幂零矩阵.

2.1.17 定理 设 $A \in \mathbb{C}^{n\times n}$ 对某个正整数 $k > n$, $A^k = 0$,则

(1) A 的特征多项式 $p_A(\lambda) = \lambda^n$.

(2) 存在正整数 $r \leq n$, $A^r = 0$.

证 (1) 设 $\tilde{\lambda}$ 是 A 的任一特征值,其相应的特征向量为 $x \in \mathbb{C}^n$, $x \neq 0$,则从 $Ax = \tilde{\lambda}x$ 及 $A^k = 0$,有

$$0 = A^k x = \tilde{\lambda}^k x,$$

推出 $\tilde{\lambda} = 0$.因此,A 的所有特征值为零,其特征多项式

$$p_A(\lambda) \equiv \det(\lambda I - A) = \lambda^n.$$

(2) 依 **1.6.9**,

$$p_A(A) = A^n = 0,$$

而且,从 A 的 Jordan 标准形(见 **5.1.2**)可知,取 r 为 A 的最大 Jordan 块

的阶,则$1 \le r \le n$,且有$A^r = 0$. □

2.1.18 定理 设$A \in \mathbb{C}^{n\times n}$,则$A$可以表示成

$$A = A_D + A_N, \tag{1.24}$$

其中$A_D \in \mathbb{C}^{n\times n}$是可通过相似变换而对角化的矩阵,$A_N \in \mathbb{C}^{n\times n}$是幂零矩阵.而且,$A_D$和$A_N$关于矩阵乘法是可交换的,即

$$A_D A_N = A_N A_D.$$

证明留作练习.提示:利用A的 Jordan 标准形(见 **5.1.2**),J可以表示成$J = D + N$,其中D是对角分块纯量矩阵,N是和D匹配的对角分块幂零矩阵.

2.1.19 定义 设$A \in \mathbb{C}^{n\times n}$.用$\lambda(A)$表示$A$的特征值全体所成之集合:

$$\lambda(A) \equiv \{\lambda \in \mathbb{C}: Ax = \lambda x, x \in \mathbb{C}^n, x \ne 0\}, \tag{1.25}$$

用$\rho(A)$表示特征值之模的最大值:

$$\rho(A) \equiv \{|\lambda|: \lambda \in \lambda(A)\}, \tag{1.26}$$

称其为A的**谱半径**.

几何上,如果把A的所有特征值描绘在复平面上,则它们包含在以原点为中心以$\rho(A)$为半径的圆盘中:

$$\lambda(A) \subset \{ z \in \mathbb{C}: |z| \le \rho(A) \},$$

而且$\rho(A)$是包含$\lambda(A)$的形如$|z| \le R$的最小圆盘的半径.

2.1.20 定理 设$\|\cdot\|$是$\mathbb{C}^{n\times n}$上任一范数,则

$$\rho(A) \le \|A\|, \quad \forall A \in \mathbb{C}^{n\times n}. \tag{1.27}$$

证 设λ是A的特征值,

$$|\lambda| = \rho(A),$$

$x \in \mathbb{C}^n$ 是相应的特征向量:

$$Ax = \lambda x .$$

取单位向量 $e_1 \equiv (1,0,\cdots,0)^{\mathrm{T}}$,由于 $x \neq 0$, xe_1^{T} 是第1列为 x 的 $n \times n$ 非零矩阵.利用矩阵范数的齐次性和相容性,

$$\rho(A)\|xe_1^{\mathrm{T}}\| \le \|\lambda xe_1^{\mathrm{T}}\| = \|Axe_1^{\mathrm{T}}\| \le \|A\|\|xe_1^{\mathrm{T}}\|,$$

由此并注意到 $\|xe_1^{\mathrm{T}}\| > 0$,即得(1.27). □

2.1.21 推论 设 $A = [a_{ij}] \in \mathbb{C}^{n\times n}$,则

$$\rho(A) \le \min\left\{\max_{1\le i\le n}\sum_{j=1}^{n}|a_{ij}|,\quad \max_{1\le j\le n}\sum_{i=1}^{n}|a_{ij}|\right\}. \tag{1.28}$$

证 在 **2.1.20** 中取 ∞ – 范数 $\|\bullet\|_{\infty}$,并应用于 A 和 A^{T}.注意 A^{T} 和 A 有相同的特征值. □

2.1.22 定理 设 $A \in \mathbb{C}^{n\times n}$,则对于任意的 $\varepsilon > 0$,存在 $\mathbb{C}^{n\times n}$ 上的算子范数 $\|\bullet\|$,使得

$$\|A\| \le \rho(A) + \varepsilon . \tag{1.29}$$

证 依 Schur 分解定理(见 **5.3.3**),存在 $n \times n$ 的酉矩阵 U (即满足条件 $U^* = U^{-1}$),使得

$$U^*AU = \Lambda + T ,$$

其中

$$\Lambda \equiv \mathrm{diag}(\lambda_1,\cdots,\lambda_n), \quad \{\lambda_1,\cdots,\lambda_n\} = \lambda(A),$$

而 $T = [t_{ij}]$ 是严格上三角矩阵.

考虑

$$D_\delta \equiv \mathrm{diag}(1,\delta,\delta^2,\cdots,\delta^{n-1}),$$

得

$$D_\delta^{-1}U^*AUD_\delta = \Lambda + D_\delta^{-1}TD_\delta ,$$

其中

$$D_\delta^{-1}TD_\delta = \begin{bmatrix} 0 & \delta t_{12} & \delta^2 t_{13} & \cdots & \delta^{n-1}t_{1n} \\ 0 & 0 & \delta t_{23} & \cdots & \delta^{n-2}t_{2n} \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ 0 & 0 & 0 & \cdots & \delta t_{n-1,n} \\ 0 & 0 & 0 & \cdots & 0 \end{bmatrix}.$$

现在对任意给定的$\varepsilon > 0$,取$\delta > 0$,使得对矩阵$1-$范数$\|\bullet\|_1$成立

$$\left\|D_\delta^{-1}TD_\delta\right\|_1 \le \varepsilon .$$

再利用向量$1-$范数$\|\bullet\|_1$定义向量范数$\|\bullet\|_{UD_\delta}$:

$$\|x\|_{UD_\delta} \equiv \left\|(UD_\delta)^{-1}x\right\|_1, \quad \forall x \in \mathbf{C}^n .$$

于是,对于由其导出的矩阵范数,有

$$\begin{aligned}\|A\| &= \sup_{x\neq 0}\frac{\|Ax\|_{UD_\delta}}{\|x\|_{UD_\delta}} = \sup_{x\neq 0}\frac{\left\|(UD_\delta)^{-1}Ax\right\|_1}{\left\|(UD_\delta)^{-1}x\right\|_1} \\ &= \sup_{x\neq 0}\frac{\left\|D_\delta^{-1}U^*AUD_\delta(UD_\delta)^{-1}x\right\|_1}{\left\|(UD_\delta)^{-1}x\right\|_1} = \left\|D_\delta^{-1}U^*AUD_\delta\right\|_1 \\ &\le \|\Lambda\|_1 + \left\|D_\delta^{-1}TD_\delta\right\| \le \rho(A) + \varepsilon .\end{aligned}$$

□

2.1.23 定理 设$A \in \mathbf{C}^{n\times n}$,则成立

$$\lim_{k\to\infty}A^k = 0 \Leftrightarrow \rho(A) < 1. \tag{1.30}$$

证 如果$\lim_{k\to\infty}A^k = 0$,假定

$$\lambda \in \lambda(A), \quad |\lambda| = \rho(A),$$

那么根据(1.27),

$$\rho(A)^k = |\lambda|^k \le \rho\left(A^k\right) \le \left\|A^k\right\|_2, \quad k = 1,2,\cdots.$$

由此,$\lim\limits_{k\to\infty}\rho(A)^k=0$,推出 $\rho(A)<1$.

反之,设 $\rho(A)<1$,根据(1.29),必有算子范数 $\|\bullet\|$ 使得 $\|A\|<1$,注意到

$$0\le\left\|A^k\right\|\le\|A\|^k,$$

即得 $\lim\limits_{k\to\infty}A^k=0$. □

2.1.24 定义 设 $A\in\mathbb{C}^{n\times n}$,如果

$$\lim_{k\to\infty}A^k=0, \tag{1.31}$$

则称 A 是**收敛矩阵**.否则,称 A 是**发散矩阵**.

2.1.25 定理 设 $A\in\mathbb{C}^{n\times n}$, $\varepsilon>0$,则存在常数 $c\equiv c(A,\varepsilon)$,使得

$$\left|\left(A^k\right)_{ij}\right|\le c(\rho(A)+\varepsilon)^k,\quad \forall k=1,2,\cdots,\quad i,j=1,\cdots,n. \tag{1.32}$$

证 令 $\tilde{A}\equiv(\rho(A)+\varepsilon)^{-1}A$,显然,

$$\rho(\tilde{A})=(\rho(A)+\varepsilon)^{-1}\rho(A)<1.$$

根据(1.30),当 $k\to\infty$ 时 $\tilde{A}^k\to 0$.从而,矩阵序列 $\left\{\tilde{A}^k\right\}$ 的各个对应元素组成的序列是一致有界的,即存在仅依赖于 $\tilde{A}$(亦即 A 和 ε)的常数 c,使得

$$\left|\left(\tilde{A}^k\right)_{ij}\right|\le c,\quad \forall k=1,2,\cdots,\quad i,j=1,\cdots,n.$$

这等价于(1.32). □

2.1.26 定理 设 $\|\bullet\|$ 是 $\mathbb{C}^{n\times n}$ 上的一种范数,则

$$\rho(A)=\lim_{k\to\infty}\left\|A^k\right\|^{1/k},\quad \forall A\in\mathbb{C}^{n\times n}. \tag{1.33}$$

证 因为

$$\rho(A)^k = \rho(A^k) \le \|A^k\|,$$

所以

$$\rho(A) \le \|A^k\|^{1/k}, \quad \forall k = 1,2,\cdots. \tag{1.34}$$

对于任意取定的$\varepsilon > 0$,令

$$\tilde{A} \equiv (\rho(A)+\varepsilon)^{-1} A.$$

由于$\rho(\tilde{A})<1$,当$k \to \infty$时有$\|\tilde{A}^k\| \to 0$,从而存在某个正整数$N \equiv N(A,\varepsilon)$,使得

$$\|\tilde{A}^k\| < 1, \quad \forall k \ge N.$$

于是,注意到

$$\|\tilde{A}^k\| < 1 \Leftrightarrow \|A^k\| \le (\rho(A)+\varepsilon)^k,$$

以及(1.34),有

$$\rho(A) \le \|A^k\|^{1/k} \le \rho(A)+\varepsilon, \quad \forall k \ge N.$$

此不等式,ε是任意的,因而等价于(1.33). □

2.1.27 定理 设$A=[a_{ij}] \in \mathbb{C}^{n\times n}$,则成立

$$\rho(A) \le \min_{p_1,\cdots,p_n>0} \max_{1\le i\le n} \frac{1}{p_i} \sum_{j=1}^{n} p_j |a_{ij}| \tag{1.35}$$

和

$$\rho(A) \le \min_{p_1,\cdots,p_n>0} \max_{1\le j\le n} p_j \sum_{i=1}^{n} \frac{1}{p_i} |a_{ij}|. \tag{1.36}$$

证 从 **1.6.25** 直接推出. □

2.1.28 定义 设$A\in\mathbb{C}^{n\times n}$,$\lambda \in \lambda(A)$.因$\lambda(A^{\mathrm{T}})=\lambda(A)$,故对$A^{\mathrm{T}}$存在属于$\lambda$的特征向量$y \in \mathbb{C}^n$:

$$A^{\mathrm{T}} y = \lambda y.$$

此式也可以写成

$$y^{\mathrm{T}}A=\lambda y^{\mathrm{T}} \tag{1.37}$$

因而称 y 为 A 的**左特征向量**.相对地,A 本身属于 λ 的特征向量称为 A 的**右特征向量**.

2.1.29 定义 设 $A\in\mathbb{C}^{n\times n}$ 有 r 个不同的特征值 $\lambda_1,\cdots,\lambda_r$.

如果 $\lambda_1,\cdots,\lambda_r$ 分别是特征多项式 $p_A(\lambda)$ 的 $n(\lambda_1),\cdots,n(\lambda_r)$ 重根:

$$p(\lambda)\equiv\det(\lambda I-A)=\prod_{i=1}^{r}(\lambda-\lambda_i)^{n(\lambda_i)},\quad \sum_{i=1}^{r}n(\lambda_i)=n, \tag{1.38}$$

则称 $n(\lambda_i)$ 为 λ_i 的**代数重数**.

A 的属于 λ_i 的线性无关特征向量的个数,记作 $m(\lambda_i)$,

$$m(\lambda_i)\equiv n-\operatorname{rank}(\lambda_i I-A) \tag{1.39}$$

称其为 λ_i 的**几何重数**.

显然,代数重数和几何重数关系如下:

$$1\le m(\lambda_i)\le n(\lambda_i)\le n,\quad i=1,\cdots,r. \tag{1.40}$$

通常,"重数"一词多指代数重数.(简)**单特征值**是指代数重数为1(从而几何重数也必为1)的特征值.

2.1.30 定理 设 $A\in\mathbb{C}^{n\times n}$,$\lambda\in\lambda(A)$.$\lambda$ 是 A 的单特征值的充分必要条件是

(1) λ 的几何重数为1,即 $\operatorname{rank}(\lambda I-A)=n-1$.

(2) 属于 λ 的右特征向量 u 和左特征向量 v 满足 $v^{\mathrm{T}}u\ne 0$.

证 相似变换不会改变特征值及其代数重数和几何重数,也不会改变右特征向量和左特征向量是否满足条件(2).因此,不妨直接

假定 A 是Jordan标准形(见**5.1.2**).这样,必要性的证明是明显的.下面给出充分性的证明.

因为 λ 的几何重数为1,可假设 A 形如

$$A=\begin{bmatrix} J_1 & 0 \\ 0 & J_2 \end{bmatrix},$$

其中 J_1 是关于 λ 的 $m\times m$ 的 Jordan 块,

$$J_1=\begin{bmatrix} \lambda & 1 & 0 & \cdots & 0 \\ 0 & \ddots & \ddots & & \vdots \\ \vdots & \ddots & \ddots & \ddots & 0 \\ \vdots & & \ddots & \ddots & 1 \\ 0 & \cdots & \cdots & 0 & \lambda \end{bmatrix},$$

而 J_2 是所有不属于 λ 的各 Jordan 块的直和.显然,在如此情形下,单位向量 e_m 和 e_1 分别是 A 的属于 λ 的左特征向量和右特征向量.如果 $m>1$,那么 $e_m^{\mathrm{T}}e_1=0$,这与条件(2)矛盾.因此,必有 $m=1$,这就是说,λ 的代数重数为1. □

2.2 不可约矩阵和对角优势矩阵

2.2.1 定义 设 $A\in\mathbb{C}^{n\times n}$,$n\geq 2$,如果存在 $n\times n$ 置换矩阵(见**1.5.14**)P,使得

$$PAP^{\mathrm{T}}=\begin{bmatrix} A_{11} & A_{12} \\ 0 & A_{22} \end{bmatrix}, \tag{2.1}$$

其中 A_{11} 是 $r\times r$ 子矩阵,A_{22} 是 $(n-r)\times(n-r)$ 子矩阵,$1\leq r\leq n$,则称 A 为**可约矩阵**(reducible matrix)或**可分矩阵**(decomposable matrix).

否则,即 A 不是可约的,称 A 为**不可约矩阵**(irreducible matrix)或**不可分矩阵**(indecomposable matrix).

"不可约"一词是 Frobenius(1912)提出的.明显的原因是对于形如(2.1)的矩阵常可约化为两个低阶矩阵来处理.例如,求解以其为系数矩阵的线性方程组,可以归结为先求解以 A_{22} 为系数矩阵的 $n-r$ 阶线性方程组,然后求解以 A_{11} 为系数矩阵的 r 阶线性方程组.

2.2.2 定义 设 $A=[a_{ij}]\in\mathbb{C}^{n\times n}$,如果成立

$$|a_{ii}|\geq\sum_{j=1,j\neq i}^{n}|a_{ij}|,\quad i=1,\cdots,n,\tag{2.2}$$

则称 A 为**行对角优势矩阵**(row diagonally dominant matrix).

如果(2.2)成立且至少对一个 i 取严格不等号,则称 A 为**行弱严格对角优势矩阵**.

如果(2.2)对每个 i 取严格不等号,则称 A 为**行严格对角优势矩阵**(row strictly diagonally dominant matrix).

如果 A^{T} 是行对角优势的,则称 A 为**列对角优势矩阵**(column diagonally dominant matrix).

如果 A^{T} 是行弱严格对角优势的,则称 A 为**列弱严格对角优势矩阵**.

如果 A^{T} 是行严格对角优势的,则称 A 为**列严格对角优势矩阵**(column strictly diagonally dominant matrix).

2.2.3 定理(Taussky) 设 $A=[a_{ij}]\in\mathbb{C}^{n\times n}$ 是(行或列)严格对角优势矩阵或不可约(行或列)弱严格对角优势矩阵,则

(1) A 是非奇异的.

(2) 如果 $a_{ii}>0,i=1,\cdots,n$,那么

$$\mathrm{Re}(\lambda_i)>0,\quad\forall\lambda_i\in\lambda(A).\tag{2.3}$$

或者说，A 是正稳定的(见 **3.14.3**).特别，当 A 的所有特征值为实数时，必全为正数：

$$\lambda_i > 0, \quad \forall \lambda_i \in \lambda(A).$$

证 考虑 A 是行严格对角优势矩阵.此时，复平面上 n 个圆盘

$$|z - a_{ii}| \le \sum_{j=1, j\ne i}^{n} |a_{ij}| \equiv \Lambda_i, \quad i = 1, \cdots, n \tag{2.4}$$

的并集必不包含原点 $z = 0$. 依 **1.6.23**, $\lambda = 0$ 不是 A 的特征值，因此，A 是非奇异的.

如果再附加条件

$$a_{ii} > 0, \quad i = 1, \cdots, n,$$

那么对任一实部 ≤ 0 的复数 c 成立

$$|c - a_{ii}| = \left[(a_{ii} - \mathrm{Re}\, c)^2 + (\mathrm{Im}\, c)^2 \right]^{1/2} \ge a_{ii} > \Lambda_i, \quad i = 1, \cdots, n.$$

这就是说，c 不属于(2.4)的 n 个圆盘的并集，因而不可能是 A 的特征值.

不可约弱严格对角优势情形的证明留作练习. □

2.2.4 定理 设 $A = [a_{ij}] \in \mathbb{R}^{n\times n}$ 是(行或列)严格对角优势矩阵，则 $\det A$ 和 A 的主对角元素之积

$$a_{11} a_{22} \cdots a_{nn}$$

同号.而且，当 A 是行严格对角优势时，

$$|\det A| \ge \prod_{i=1}^{n} \left(|a_{ii}| - \sum_{j\ne i} |a_{ij}| \right). \tag{2.5}$$

当 A 是列严格对角优势时，

$$|\det A| \ge \prod_{j=1}^{n} \left(|a_{jj}| - \sum_{i\ne j} |a_{ij}| \right). \tag{2.6}$$

证 由假设知

$$a_{ii} \neq 0, i=1,\cdots,n.$$

记

$$\varepsilon_i = \operatorname{sgn} a_{ii}, \quad i=1,\cdots,n; \quad B \equiv [\varepsilon_i a_{ij}] \in \mathbb{R}^{n\times n}, \quad (2.7)$$

于是

$$\det A = \varepsilon_1 \varepsilon_2 \cdots \varepsilon_n \det B. \quad (2.8)$$

注意到 B 的对角元素是正实数:

$$\varepsilon_i a_{ii} = |a_{ii}| > 0, \quad i=1,\cdots,n.$$

依 **2.2.3**, B 的所有特征值具有正实部.这样,由于 B 是实的,复特征值必共轭成对出现,其积是正实数,而实特征值则必为正实数,从而 $\det B$——等于 B 的所有特征值(按代数重数计)之积——必大于零.因此从(2.8)有

$$\operatorname{sgn}(\det A) = \varepsilon_1 \varepsilon_2 \cdots \varepsilon_n = \operatorname{sgn}(a_{11} a_{22} \cdots a_{nn}).$$

不等式(2.5)和(2.6)的证明从略,可作练习. □

2.2.5 定义 设 $A=[a_{ij}] \in \mathbb{C}^{n\times n}$,如果成立

$$|a_{ii}| > |a_{ij}|, \quad j=1,\cdots,n, j\neq i; \quad i=1,\cdots,n, \quad (2.9)$$

则称 A 是**行元素严格对角优势矩阵**.如果成立

$$|a_{ii}| > |a_{ji}|, \quad j=1,\cdots,n, j\neq i; \quad i=1,\cdots,n, \quad (2.10)$$

则称 A 是**列元素严格对角优势矩阵**.

显然,行严格对角优势矩阵必是行元素严格对角优势矩阵,反之则不然.

2.2.6 定理 设 $A \in \mathbb{R}^{n\times n}$ 是行严格对角优势矩阵,则 A^{-1} 是列元素严格对角优势矩阵.

证 依 **2.2.3**, A 可逆.设

$$A=[a_{ij}], \quad A^{-1}=[\alpha_{ij}].$$

依 **1.5.13**,

$$\alpha_{ij}=(-1)^{i+j}\frac{\det A_{ij}}{\det A},\quad i,j=1,\cdots,n.$$

因此,只须证明

$$|\det A_{ii}|>|\det A_{ij}|,\quad i=1,\cdots,n,\ i\neq j.$$

不失一般性,为了方便,取

$$i=1,\quad j=2,$$

利用(2.7)中的记号,从 **2.2.4** 的证明得知

$$\det B_{11}>0.$$

注意到

$$\det B_{11}=|\det A_{11}|,\ |\det B_{12}|=|\det A_{12}|.$$

为了完成定理的证明,只须证明

$$\det B_{11}\pm\det B_{12}>0.$$

事实上,

$$\det B_{11}\pm\det B_{12}=\det\begin{bmatrix}\varepsilon_2a_{22} & \varepsilon_2a_{23} & \cdots & \varepsilon_2a_{2n}\\ \varepsilon_3a_{32} & \varepsilon_3a_{33} & \cdots & \varepsilon_3a_{3n}\\ \cdots & \cdots & \cdots & \cdots\\ \varepsilon_na_{n2} & \varepsilon_na_{n3} & \cdots & \varepsilon_na_{nn}\end{bmatrix}$$

$$\pm\det\begin{bmatrix}\varepsilon_2a_{21} & \varepsilon_2a_{23} & \cdots & \varepsilon_2a_{2n}\\ \varepsilon_3a_{31} & \varepsilon_3a_{33} & \cdots & \varepsilon_3a_{3n}\\ \cdots & \cdots & \cdots & \cdots\\ \varepsilon_na_{n1} & \varepsilon_na_{n3} & \cdots & \varepsilon_na_{nn}\end{bmatrix}$$

$$=\det\begin{bmatrix}\varepsilon_2(a_{22}\pm a_{21}) & \varepsilon_2a_{23} & \cdots & \varepsilon_2a_{2n}\\ \varepsilon_3(a_{32}\pm a_{31}) & \varepsilon_3a_{33} & \cdots & \varepsilon_3a_{3n}\\ \cdots & \cdots & \cdots & \cdots\\ \varepsilon_n(a_{n2}\pm a_{n1}) & \varepsilon_na_{n3} & \cdots & \varepsilon_na_{nn}\end{bmatrix}.$$

此式中最后的行列式是正的,因为其矩阵是行严格对角优势且对角

元素全大于零. □

2.2.7 定义 矩阵 $B=[b_{ij}]\in\mathbb{C}^{n\times n}$ 称为**等模**(equimodular)于矩阵 $A=[a_{ij}]\in\mathbb{C}^{n\times n}$,如果

$$|b_{ij}|=|a_{ij}|,\quad \forall i,j=1,\cdots,n. \tag{2.11}$$

2.2.8 定理 设 $A\in\mathbb{C}^{n\times n}$,等模于 A 的每个矩阵均非奇异的充分必要条件是 A 可以表示成

$$A=PMD, \tag{2.12}$$

其中 $P\in\mathbb{R}^{n\times n}$ 是置换矩阵,$D\in\mathbb{R}^{n\times n}$ 是对角元素全为正实数的对角矩阵,$M\in\mathbb{C}^{n\times n}$ 是行严格对角优势矩阵.

此定理证明从略.可参见[CH1966].

2.2.9 定理 设 $A=[a_{ij}]\in\mathbb{C}^{n\times n}$ 行严格对角优势矩阵,则对于任何 $B=[b_{ij}]\in\mathbb{C}^{n\times n}$ 成立

$$\left\|A^{-1}B\right\|_\infty\le\max_{1\le i\le n}\frac{\sum_{j=1}^{n}\left|b_{ij}\right|}{\left|a_{ii}\right|-\sum_{j=1,j\ne i}^{n}\left|a_{ij}\right|}. \tag{2.13}$$

证 依算子范数定义 **1.8.30**,存在

$$\xi=\left(\xi_1,\cdots,\xi_n\right)^{\mathrm T}\in\mathbb{C}^n,\ \ \|\xi\|_\infty=1,$$

使得

$$\left\|A^{-1}B\right\|_\infty=\left\|A^{-1}B\xi\right\|_\infty.$$

令

$$\eta=\left(\eta_1,\cdots,\eta_n\right)^{\mathrm T}=A^{-1}B\xi,$$

且令

$$\left|\eta_{i_0}\right|=\max_{1\le i\le n}\left|\eta_i\right|=\|\eta\|_\infty,$$

由 $A\eta = B\xi$ 得

$$\sum_{j=1}^{n} a_{i_0 j}\eta_j = \sum_{j=1}^{n} b_{i_0 j}\xi_j .$$

于是

$$\begin{aligned}|\eta_{i_0}|\left(|a_{i_0 i_0}| - \sum_{j=1,j\neq i_0}^{n} |a_{i_0 j}|\right) &\le |a_{i_0 i_0}\eta_{i_0}| - \sum_{j=1,j\neq i_0}^{n} |a_{i_0 j}\eta_j| \\ &\le \left|\sum_{j=1}^{n} a_{i_0 j}\eta_j\right| = \left|\sum_{j=1}^{n} b_{i_0 j}\xi_j\right| \le \sum_{j=1}^{n}|b_{i_0 j}\xi_j| \le \sum_{j=1}^{n}|b_{i_0 j}|,\end{aligned}$$

从而

$$\begin{aligned}\|A^{-1}B\|_\infty &= \|\eta\|_\infty = |\eta_{i_0}| \\ &\le \frac{\sum_{j=1}^{n}|b_{i_0 j}|}{|a_{i_0 i_0}| - \sum_{j=1,j\neq i_0}^{n}|a_{i_0 j}|} \le \max_{1\le i\le n}\frac{\sum_{j=1}^{n}|b_{ij}|}{|a_{ii}| - \sum_{j=1,j\neq i}^{n}|a_{ij}|}.\end{aligned}$$

□

2.2.10 定理 设 $A=[a_{ij}]\in\mathbb{C}^{n\times n}$ 是所有对角元素不为零的对角优势矩阵,并且对于 $i=1,\cdots,n$,成立

$$\sum_{j=1,j\neq i}^{n}|a_{ij}| < |a_{ii}| \quad i\neq k;\quad \sum_{j=1,j\neq k}^{n}|a_{kj}| = |a_{kk}|, \qquad (2.14)$$

则 A 是可逆的.

证 取

$$p_i = 1, i=1,\cdots,n, i\neq k;\ p_k = 1+\varepsilon, \varepsilon>0,$$

则有

$$\frac{1}{p_k}\sum_{j=1,j\neq k}^{n} p_j|a_{kj}| = \frac{1}{1+\varepsilon}\sum_{j=1,j\neq k}^{n}|a_{kj}| < |a_{kk}|$$

和

$$\frac{1}{p_i}\sum_{j=1,j\neq i}^{n} p_j|a_{ij}| = \sum_{j=1,j\neq k}^{n}|a_{ij}| + \varepsilon|a_{ik}|,\ i=1,\cdots,n,i\neq k.$$

利用(2.14),可以选取足够小的 $\varepsilon>0$,使得

$$\sum_{j=1,j\neq k}^{n}|a_{ij}| + \varepsilon|a_{ik}| < |a_{ii}|,\ i=1,\cdots,n,i\neq k.$$

于是,依 **1.6.25**,点 $z=0$ 必不在 Gerschgorin 圆内.因此,A 必是可逆的. □

2.2.11 定理 设 $A=[a_{ij}]\in\mathbb{C}^{n\times n}$ 是弱严格对角优势矩阵,而且其所有元素不为零,则 A 是可逆的.

证 若 A 不可逆,则 0 是 A 的一个特征值.因 A 有弱严格对角优势,0 不可能是 A 的 Gerschgorin 域 $G(A)$ 的内点,故必为 $G(A)$ 的边界点.于是,依 **1.6.27**, A 的每个 Gerschgorin 圆通过 0.但是,由于至少有一个指标 i,成立

$$|a_{ii}| > \sum_{j=1,j\neq i}^{n}|a_{ij}|,$$

表明第 i 个 Gerschgorin 圆不可能通过 0.矛盾. □

2.3 酉矩阵和实正交矩阵

2.3.1 引言 **1.7.29** 已经建立酉映射的概念.现在平行地引入酉矩阵的定义,并介绍其一些最基本的性质;其进一步的性质,将陆续在以后的有关章节中论述.

2.3.2 定义 $U\in\mathbb{C}^{n\times n}$ 称为**酉矩阵**(unitary matrix),如果

$$U^*U=I. \tag{3.1}$$

当$U \in \mathbb{R}^{n\times n}$时,(3.1)成为

$$U^{\mathrm{T}}U = I. \tag{3.2}$$

U 即为**实正交矩阵**(见 **1.7.26**).

2.3.3 定理 设$U \in \mathbb{C}^{n\times n}$,下列诸条件是等价的:

(1) U 是酉矩阵.

(2) U 是非奇异的,而且$U^* = U^{-1}$.

(3) $UU^* = I$.

(4) U^*是酉矩阵.

(5) U 的各列是单位向量且两两正交.

(6) U 的各行是单位向量且两两正交.

(7) 若$x \in \mathbb{C}^n$, $y = Ux$,则$\|y\|_2 = \|x\|_2$.

(8) $(Ux, Uy) = (x, y), \quad \forall x, y \in \mathbb{C}^n$.

证 条件$U^*U = I$等价于

$$|\det A| = 1, \quad U^* = U^{-1}.$$

如此条件等价于

$$UU^* = I,$$

这一条件可以写成

$$(U^*)^* U^* = I,$$

故又等价于U^*是酉矩阵.因此(1)$\Leftrightarrow$(2)$\Leftrightarrow$(3)$\Leftrightarrow$(4).

现设$\boldsymbol{u}^{(1)}, \cdots, \boldsymbol{u}^{(n)}$是$U$ 的n 个列,条件$U^*U = I$的另一形式是

$$(\boldsymbol{u}^{(i)})^* \boldsymbol{u}^{(j)} = \begin{cases} 0, & j \neq i, \\ 1, & j = i, \end{cases} \quad i, j = 1, \cdots, n,$$

因此(1)$\Leftrightarrow$(5).

类似地,(4)$\Leftrightarrow$(6).

如果$U^*U = I$且$y = Ux$,那么

$$\|y\|_2^2 = y^*y = x^*U^*Ux = x^*Ix = x^*x = \|x\|_2^2,$$

所以(1)$\Rightarrow$(7).

下面证明(7)$\Rightarrow$(1).

设成立条件(7).首先考虑$n=2$的情形.

先取$x=(1,0)^{\mathrm{T}}$,得U^*U 的(1,1)-元素

$$(U^*U)_{1,1}=x^*U^*Ux=y^*y=x^*x=1.$$

再取 $x=(0,1)^{\mathrm{T}}$,类似地,得U^*U 的(2,2)-元素也是1.于是,U^*U 必定形如

$$\begin{bmatrix}1 & a\\ \overline{a} & 1\end{bmatrix},$$

其中a是U的第1列和第2列的内积,$\overline{a}$是U的第2列和第1列的内积.

至此,若取$x=(1,1)^{\mathrm{T}}$,则

$$2=x^*x=y^*y=x^*U^*Ux=2+(a+\overline{a}).$$

再取$x=(1,i)^{\mathrm{T}}$,$\mathrm{i}=\sqrt{-1}$,类似地,得

$$2=2+\mathrm{i}(a-\overline{a}),$$

于是有

$$2\operatorname{Re}a=a+\overline{a}=0$$

和

$$2i\operatorname{Im}a=a-\overline{a}=0,$$

推出$a=0$.

因此,对于$U\in\mathbb{C}^{2\times2}$的情形来说,如果成立

$$x^*U^*Ux=x^*x,\quad \forall x\in\mathbb{C}^2,$$

则必有$U^*U=I$,即U是酉矩阵.

现在考虑$n>2$的情形.令

$$A=[a_{ij}]\equiv U^*U,$$

取

$$x=(0,\cdots,0,x_i,0,\cdots,0,x_j,0,\cdots,0)^{\mathrm{T}}\in\mathbb{C}^n,$$

则

$$x^*Ax = (\overline{x_i}, \overline{x_j})\begin{bmatrix} a_{ii} & a_{ij} \\ a_{ji} & a_{jj} \end{bmatrix}\begin{pmatrix} x_i \\ x_j \end{pmatrix}.$$

对此,已证明条件(7)蕴涵

$$\begin{bmatrix} a_{ii} & a_{ij} \\ a_{ji} & a_{jj} \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}.$$

因此,由于 i 和 j 是任意的, A 的每个 2×2 主子矩阵均为 2×2 单位矩阵.显然,如此矩阵 A 只能是 $A = I$.注意, $n = 1$ 时(7)明显地蕴涵(1).

最后,证明(1) $\Leftrightarrow$ (8).

若 $U^*U = I$,则

$$(Ux, Uy) = (U^*Ux, y) = (x, y), \quad \forall x, y \in \mathbb{C}^n.$$

反之,若条件(8)成立,则

$$(U^*Ux, y) = (x, y),$$

于是

$$((U^*U - I)x, y) = 0, \quad \forall x, y \in \mathbb{C}^n.$$

特别,取 $y = (U^*U - I)x$,得

$$\|(U^*U - I)x\| = 0, \quad \forall x \in \mathbb{C}^n,$$

从而

$$(U^*U - I)x = 0, \quad \forall x \in \mathbb{C}^n,$$

这等价于

$$U^*U - I = 0,$$

即 U 是酉矩阵. □

2.3.4 定理 设 $U \in \mathbb{C}^{n\times n}$ 是酉矩阵,则

$$|\lambda| = 1, \quad \lambda \in \lambda(U).$$

证 设 $\lambda \in \lambda(U)$, $x \in \mathbb{C}^n$ 是相应的特征向量.依 **2.3.3** 的酉矩阵的等价条件(7),

$$(x,x)=(Ux,Ux)=(\lambda x,\lambda x)=|\lambda|^2(x,x),$$

由此,因$(x,x)\neq 0$,推出$|\lambda|=|\lambda|^2=1$. □

2.3.5 定义 $M\in\mathbb{C}^{n\times n}$称为**等距矩阵**,如果线性变换

$$y=Mx,\quad \forall x\in\mathbb{C}^n$$

保持 Euclid 长度不变:$\|Mx\|_2=\|x\|_2, x\in\mathbb{C}^n$.

2.3.3 表明,M 为等距矩阵的充分必要条件是M 为酉矩阵.

这里,等距矩阵平行于 **1.7.24** 定义的等距映射.

在赋予不同距离结构的各种距离空间(或称度量空间)中,有不同的"长度",分别地可以确立不同的等距性.

2.3.6 定理 设$U,V\in\mathbb{C}^{n\times n}$是酉矩阵(实正交矩阵),则积$UV$ 也是酉矩阵(实正交矩阵).

证 $(UV)^*(UV)=V^*(U^*U)V=V^*V=I$. □

2.3.7 定理 设$\{U^{(k)}\}\subset\mathbb{C}^{n\times n}$是酉矩阵序列,则存在收敛子序列,且极限矩阵也是酉矩阵.

证 酉矩阵的每列的 Euclid 长度等于1,其每个元素的绝对值不会大于1.设

$$U^{(k)}\equiv\left[u_{ij}^{(k)}\right],\quad k=1,2,\cdots$$

考虑其(1,1)-元素构成的复数序列

$$u_{11}^{(1)},u_{11}^{(2)},\cdots,u_{11}^{(k)},\cdots \tag{3.3}$$

这是一个有界序列:

$$\left|u_{11}^{(k)}\right|\leq 1,\quad k=1,2,\cdots$$

根据分析数学基本原理——在紧集上的无穷序列必有收敛子序列,(3.3)有收敛子序列:

$$u_{11}^{(k_{11})}, u_{11}^{(k_{12})}, \cdots, u_{11}^{(k_{1i})}, \cdots, \quad \lim_{i\to\infty} u_{11}^{(k_{1i})} = u_{11}^{(0)}.$$

现在,对于 $\{U^{(k)}\}$ 的子序列 $\{U^{(k_{1i})}\}$,考虑其 $(1,2)$ – 元素构成的复数序列

$$u_{12}^{(k_{11})}, u_{12}^{(k_{12})}, \cdots, u_{12}^{(k_{1i})}, \cdots$$

同样,这一序列有收敛子序列

$$u_{12}^{(k_{21})}, u_{12}^{(k_{22})}, \cdots, u_{12}^{(k_{2i})}, \cdots, \quad \lim_{i\to\infty} u_{12}^{(k_{2i})} = u_{12}^{(0)}.$$

如此继续,逐次考虑由相应元素组成的子序列.经 $n\times n$ 次选取,最后得 (n,n) – 元素构成的收敛子序列

$$u_{nn}^{(k_1)}, u_{nn}^{(k_2)}, \cdots, u_{nn}^{(k_i)}, \cdots, \quad \lim_{i\to\infty} u_{nn}^{(k_i)} = u_{nn}^{(0)}.$$

这样,得到序列 $\{U^{(k)}\}$ 的一个收敛子序列

$$\{U^{(k_i)}\}, \quad \lim_{i\to\infty} U^{(k_i)} = U^{(0)} = [u_{ij}^{(0)}].$$

由于

$$(U^{(k_i)})^* U^{(k_i)} = I, \quad i = 1,2,\cdots$$

从而有

$$\lim_{i\to\infty}(U^{(k_i)})^* U^{(k_i)} = (U^{(0)})^* U^{(0)} = I,$$

因此,$U^{(0)}$ 是酉矩阵. □

2.3.8 定理 设 $A\in\mathbf{C}^{n\times n}$ 是非奇异矩阵,A^{-1} 相似于 A^* 的充分必要条件是存在非奇异矩阵 $B\in\mathbf{C}^{n\times n}$,使得

$$A = B^{-1}B^*. \tag{3.4}$$

证 先证充分性.若 $A = B^{-1}B^*$, $B\in\mathbf{C}^{n\times n}$ 非奇异,则

$$B^* A^{-1}(B^*)^{-1} = B(B^*)^{-1} = (B^{-1}B^*)^* = A^*.$$

再证必要性.若 A^{-1} 相似于 A^*,则存在非奇异矩阵 $S\in\mathbf{C}^{n\times n}$,使得

$$SA^{-1}S^{-1} = A^*.$$

令

$$S_\theta \equiv e^{i\theta}S, \quad \theta \in \mathbb{R},$$

注意到

$$S_\theta A^{-1} S_\theta^{-1} = e^{i\theta} S A^{-1}\left(e^{-i\theta}S^{-1}\right) = SA^{-1}S^{-1} = A^*,$$

于是

$$S_\theta = A^* S_\theta A, \quad S_\theta^* = A^* S_\theta^* A.$$

将这两式相加,并令 $H_\theta \equiv S_\theta + S_\theta^*$,得

$$H_\theta = A^* H_\theta A.$$

如果 H_θ 是奇异的,必存在非零向量 $x \in \mathbb{C}^n$,使得

$$0 = H_\theta x = S_\theta x + S_\theta^* x,$$

从而

$$-x = S_\theta^{-1} S_\theta^* x = e^{-2i\theta} S^{-1} S^* x,$$

或者写成

$$S^{-1} S^* x = -e^{2i\theta} x.$$

这表明 $-e^{2i\theta}$ 是 $S^{-1}S^*$ 的特征值.因此,如果选取 $\theta_0 \in [0, 2\pi)$,使得 $-e^{2i\theta_0}$ 不是 $S^{-1}S^*$ 的特征值;那么 $H \equiv H_{\theta_0}$ 是非奇异的,而且具有性质 $H = A^* HA$.

现在选取 $\alpha \in \mathbb{C}, |\alpha| = 1$ 而且 α 不是 A^* 的特征值.令

$$B \equiv \beta\left(\alpha I - A^*\right)H,$$

其中 $\beta \in \mathbb{C}$ 是待选取的非零参数.显然,B 是非奇异的,这里的目的是使 $A = B^{-1}B^*$ 或写成 $BA = B^*$.由于

$$B^* = H\left(\overline{\beta}\overline{\alpha} I - \overline{\beta} A\right),$$

而且

$$\begin{aligned} BA &= \beta\left(\alpha I - A^*\right)HA = \beta\left(\alpha HA - A^* HA\right) \\ &= \beta\left(\alpha HA - H\right) = H\left(\alpha\beta A - \beta I\right), \end{aligned}$$

因而,欲使 $BA = B^*$,只须取

$$\beta = -\overline{\beta\alpha} \neq 0.$$

为此,如果 $\alpha = e^{i\psi}$,可取 $\beta = e^{i(\pi-\psi)/2}$. □

酉矩阵 U 具有性质 $U^{-1} = U^*$.扩展酉矩阵如此观念的途经之一是考察满足 A^{-1} 相似于 A^* 的矩阵 $A \in \mathbb{C}^{n\times n}$,上面的定理回答了此类矩阵的范围是集合

$$\left\{A \equiv B^{-1}B^* : B \in \mathbb{C}^{n\times n}, \det B \neq 0\right\}. \tag{3.5}$$

2.3.9 定理 设矩阵 $A \in \mathbb{C}^{n\times n}$ 相似于酉矩阵 $U \in \mathbb{C}^{n\times n}$,则 A^{-1} 相似于 A^*.

证 依假设,存在非奇异矩阵 $S \in \mathbb{C}^{n\times n}$,使得

$$A = S^{-1}US,$$

因此

$$\begin{aligned} A^{-1} &= \left(S^{-1}US\right)^{-1} = S^{-1}U^{-1}S = S^{-1}U^*S \\ &= \left(S^*S\right)^{-1}\left(S^{-1}US\right)^*\left(S^*S\right) = \left(S^*S\right)^{-1}A^*\left(S^*S\right). \end{aligned}$$ □

2.3.10 定义 矩阵 $B \in \mathbb{C}^{n\times n}$ 称为**酉等价**于矩阵 $A \in \mathbb{C}^{n\times n}$,如果存在酉矩阵 $U \in \mathbb{C}^{n\times n}$,使得

$$B = U^*AU. \tag{3.6}$$

如果 U 可以取成实的(因而是实正交的),则 B 称为**(实)正交等价**于 A.酉等价也称为**酉相似**.

2.3.11 定理 设 $A = [a_{ij}] \in \mathbb{C}^{n\times n}$ 和 $B = [b_{ij}] \in \mathbb{C}^{n\times n}$ 是酉等价的,则

$$\sum_{i,j=1}^{n}\left|b_{ij}\right|^2 = \sum_{i,j=}^{n}\left|a_{ij}\right|^2. \tag{3.7}$$

证 从矩阵乘法看出

$$\sum_{i,j=1}^{n}\left|a_{ij}\right|^2 = \mathrm{tr}A^*A,$$

因此,只须校验成立 $\operatorname{tr}B^*B=\operatorname{tr}A^*A$.

现设 $B=U^*AU$,注意到矩阵的迹是**相似**(变换下的)**不变量**(见**1.5.18**),即得

$$\operatorname{tr}B^*B=\operatorname{tr}U^*A^*UU^*AU=\operatorname{tr}U^*A^*AU=\operatorname{tr}A^*A. \qquad \square$$

这个定理是说,$\operatorname{tr}A^*A$ 也是矩阵 $A\in\mathbb{C}^{n\times n}$ 的**酉相似不变量**.

2.4 正规矩阵

2.4.1 定义 $A\in\mathbb{C}^{n\times n}$ 称为**正规矩阵**(normal matrix),如果

$$A^*A=AA^*. \tag{4.1}$$

显然,所有酉矩阵是正规矩阵.

2.4.2 定理 $A\in\mathbb{C}^{n\times n}$ 为正规矩阵的充分必要条件是酉等价于 A 的每个矩阵均为正规矩阵.

证 若 $B\in\mathbb{C}^{n\times n}$ 酉等价于 A,即存在酉矩阵 $U\in\mathbb{C}^{n\times n}$,使得 $B=U^*AU$,则

$$\begin{aligned}B^*B&=(U^*AU)^*(U^*AU)=U^*A^*UU^*AU\\&=U^*A^*AU=U^*AA^*U=U^*AUU^*A^*U\\&=(U^*AU)(U^*AU)^*=BB^*.\end{aligned}$$

表明 B 是正规矩阵.

反之,设酉等价于 A 的每个矩阵均为正规矩阵.显然,A 酉等价于其自身(取酉矩阵 $U=I$),故 A 是正规矩阵. $\square$

2.4.3 定义 如果 $A\in\mathbb{C}^{n\times n}$ 酉等价于对角矩阵,则称 A 为**酉可对角化矩阵**.

如果 $A \in \mathbb{C}^{n\times n}$ 正交等价于对角矩阵,则称 A 为**正交可对角化矩阵**.

2.4.4 正规矩阵谱定理　设 $A=[a_{ij}]\in \mathbb{C}^{n\times n}$, $\lambda_1,\cdots,\lambda_n \in \lambda(A)$,则下列条件等价:

(1)　A 是正规矩阵.

(2)　A 是酉可对角化矩阵.

(3)　$\sum\limits_{i,j=1}^{n}|a_{ij}|^2=\sum\limits_{i=1}^{n}|\lambda_i|^2$.

(4)　A 存在 n 个特征向量的正交组.

证　首先利用 Schur 分解定理(见 **5.3.3**), A 酉等价于一个上三角矩阵 $T=[t_{ij}]\in \mathbb{C}^{n\times n}$:

$$T=U^*AU,$$

其中 $U\in \mathbb{C}^{n\times n}$ 是某个酉矩阵. 依 **2.4.2**,条件(1)蕴涵 T 是正规矩阵.

下面证明(1)$\Leftrightarrow$(2),(2)$\Leftrightarrow$(3),(2)$\Leftrightarrow$(4).

(1)$\Leftrightarrow$(2): 若 A 是酉可对角化的,则因对角矩阵显然是正规矩阵及酉等价保持正规性, A 必为正规矩阵.

现设 A 是正规矩阵,则上三角矩阵 T 是正规矩阵:

$$T^*T=TT^*.$$

这样,可以利用这一矩阵等式两边对应位置的元素相等,(1,1)位置的元素相等:

$$\overline{t_{11}}t_{11}=t_{11}\overline{t_{11}}+\sum_{j=2}^{n}t_{1j}\overline{t_{1j}},$$

推出

$$\sum_{j=2}^{n}|t_{1j}|^2=\sum_{j=2}^{n}t_{1j}\overline{t_{1j}}=0,$$

即有

$$t_{1j}=0,\quad j=2,\cdots,n.$$

(2,2)位置的元素相等(注意$t_{12}=0$):

$$\overline{t_{22}}t_{22}=t_{22}\overline{t_{22}}+\sum_{j=3}^{n}t_{2j}\overline{t_{2j}},$$

推出$\sum_{j=3}^{n}\left|t_{2j}\right|^2=\sum_{j=3}^{n}t_{2j}\overline{t_{2j}}=0$,即

$$t_{2j}=0,\quad j=3,\cdots,n.$$

按如此方式,假定成立

$$t_{ij}=0,\quad j>i;\quad i=1,\cdots,k-1,$$

则可推出

$$t_{kj}=0,\quad j=k+1,\cdots,n,$$

最后得到

$$t_{ij}=0,\quad j>i;\quad i=1,\cdots,n.$$

因此T是对角矩阵.

$(2)\Leftrightarrow(3)$: 酉等价于A的对角矩阵的对角元素是A的特征值$\lambda_1,\cdots,\lambda_n$(按某一顺序),利用**2.3.11**,即可从(2)推出(3).另一方面,因为A的特征值$\lambda_1,\cdots,\lambda_n$是$T$的对角元素(按某一顺序),依**2.3.11**,成立

$$\sum_{i,j=1}^{n}\left|a_{ij}\right|^2=\sum_{i=1}^{n}\left|\lambda_i\right|^2+\sum_{i=1}^{n-1}\sum_{j=i+1}^{n}\left|t_{ij}\right|^2.$$

条件(3)是说

$$\sum_{i=1}^{n-1}\sum_{j=i+1}^{n}\left|t_{ij}\right|^2=0,$$

亦即T是对角矩阵.因此,(3)推出(2).

$(2)\Leftrightarrow(4)$: 设有酉矩阵U,使得$U^*AU=\Lambda$是对角矩阵,则

$$AU=U\Lambda$$

这表明Λ的对角元素是A的特征值,而U的n个列正是特征向量的正交组.

反之,如果 A 有特征向量正交组 $u^{(1)},\cdots,u^{(n)}$,记以它们为列的矩阵为 U,那么

$$
\begin{aligned}
U^*AU &= U^*\begin{bmatrix} Au^{(1)} & \cdots & Au^{(n)} \end{bmatrix} \\
&= U^*\begin{bmatrix} \lambda_1 u^{(1)} & \cdots & \lambda_n u^{(n)} \end{bmatrix} \\
&= U^*\begin{bmatrix} u^{(1)} & \cdots & u^{(n)} \end{bmatrix}\Lambda = U^*U\Lambda = \Lambda
\end{aligned}
$$

其中 λ_i 是 A 的对应特征向量 $u^{(i)}$ 的特征值,$i=1,\cdots,n$,而

$$\Lambda = \mathrm{diag}(\lambda_1,\cdots,\lambda_n). \qquad \square$$

2.4.5 定理 设 $A\in\mathbb{C}^{n\times n}$ 是正规矩阵,$\lambda\in\mathbb{C}, x\in\mathbb{C}^n$,则下列条件等价:

(1) x 是 A 属于特征值 λ 的右特征向量:

$$Ax=\lambda x.$$

(2) x 是 A^* 属于特征值 $\overline{\lambda}$ 的右特征向量:

$$A^*x=\overline{\lambda}x. \tag{4.2}$$

(3) x^* 是 A 属于特征值 λ 的左特征向量:

$$x^*A=\lambda x^*.$$

证 如果对于 A 能从(1)推出(2),那么将其应用与 A^*,直接得到

$$Ax=(A^*)^*x=\left(\overline{\overline{\lambda}}\right)x=\lambda x,$$

即从(2)推出(1).而(2)和(3)等价则基于简单关系式

$$(A^*x)^*=x^*A,\quad (\overline{\lambda}x)^*=\lambda x^*$$

和

$$(x^*A)^*=A^*x,\quad (\lambda x)^*=\overline{\lambda}x^*,$$

因此,下面只须证明(1)$\Rightarrow$(2).

不妨设 x 是规范化的,即 $x^*x=1$.依 **2.4.4**,存在酉矩阵 $U\in\mathbb{C}^{n\times n}$,其第 1 列是 x,使得

$$A=U\Lambda U^*,$$

其中 Λ 是对角矩阵,其左上角的元素是 λ.于是

$$A^* = \left(U\Lambda U^*\right)^* = \left(U^*\right)^* \Lambda^* U^* = U\overline{\Lambda}U^*,$$

从而 $A^*U = U\overline{\Lambda}$,此等式两端的第 1 列分别是 A^*x 和 $\overline{\lambda}x$,因此(4.2)成立. □

2.4.6 定理 设 $A \in \mathbb{C}^{n\times n}$ 是正规矩阵.则

(1) A^*A 的特征值全体

$$\lambda\left(A^*A\right) = \lambda\left(AA^*\right) = \left\{\ |\lambda|^2 : \lambda \in \lambda\left(A\right)\right\}. \qquad (4.3)$$

(2) A 的奇异值全体

$$\sigma\left(A\right) = \left\{|\lambda| : \lambda \in \lambda\left(A\right)\right\}. \qquad (4.4)$$

证 设 $\lambda \in \lambda\left(A\right)$, x 是 A 的属于 λ 的特征向量.依 **2.4.5**,

$$\left(A^*A\right)x = A^*\left(Ax\right) = \lambda A^*x = \lambda\overline{\lambda}x = |\lambda|^2 x,$$

因此(1)成立.由(1)及奇异值定义 **5.4.1** 即得(2). □

2.4.7 定理 正规矩阵的属于不同特征值的特征向量必相互正交.

证 设 x 和 y 分别是正规矩阵 $A \in \mathbb{C}^{n\times n}$ 的属于特征值 λ 和 μ 的特征向量,

$$Ax = \lambda x, \quad Ay = \mu y.$$

依 **2.4.5**,有

$$\mu x^*y = x^*\left(Ay\right) = \left(A^*x\right)^* y = \left(\overline{\lambda}x\right)^* y = \lambda x^* y.$$

由此,若 $\lambda \neq \mu$,则 $x^*y = 0$. □

2.4.8 定理 设 $A \in \mathbb{R}^{n\times n}$, A 为正规矩阵的充分必要条件是存在实正交矩阵 $Q \in \mathbb{R}^{n\times n}$,使得

$$Q^{\mathrm{T}}AQ = \begin{bmatrix} A_1 & 0 & \cdots & 0 \\ 0 & \ddots & \ddots & \vdots \\ \vdots & \ddots & \ddots & 0 \\ 0 & \cdots & 0 & A_k \end{bmatrix} \in \mathbb{R}^{n\times n}, \qquad (4.5)$$

其中$k \le n$,每个A_j是实1×1矩阵或形如

$$A_j = \begin{bmatrix} \alpha_j & \beta_j \\ -\beta_j & \alpha_j \end{bmatrix} \in \mathbb{R}^{2\times 2}. \tag{4.6}$$

证 充分性.对于形如(4.6)的矩阵有

$$A_j A_j^{\mathrm{T}} = \mathrm{diag}\left(\alpha_j^2 + \beta_j^2, \alpha_j^2 + \beta_j^2\right) = A_j^{\mathrm{T}} A_j .$$

表明A_j是正规的,因此(4.5)的矩阵必是正规的.依**2.4.2**, A是正规的.

必要性.应用**5.3.5**,不妨直接假定A形如

$$A = \begin{bmatrix} R & A_{01} & A_{02} & \cdots & A_{0k} \\ 0 & A_{11} & A_{12} & \cdots & A_{1k} \\ 0 & 0 & A_{22} & \cdots & A_{2k} \\ \vdots & & \ddots & \ddots & \vdots \\ 0 & 0 & \cdots & 0 & A_{kk} \end{bmatrix} \in \mathbb{R}^{n\times n}, \tag{4.7}$$

其中

$$R = \begin{bmatrix} \lambda_1 & * & \cdots & * \\ 0 & \ddots & \ddots & \vdots \\ \vdots & \ddots & \ddots & * \\ 0 & \cdots & 0 & \lambda_p \end{bmatrix} \in \mathbb{R}^{p\times p} \tag{4.8}$$

是上三角矩阵, $A_{01}, \cdots, A_{0k} \in \mathbb{R}^{p\times 2}$,并且

$$A_{ij} \in \mathbb{R}^{2\times 2}, \quad i, j = 1, \cdots, k, \quad j \ge i .$$

由于A是正规的和实的,从而成立

$$A^{\mathrm{T}} A = A A^{\mathrm{T}}$$

此等式两端的$p\times p$前主对角块相等,得

$$R^{\mathrm{T}} R = R R^{\mathrm{T}} + A_{01} A_{01}^{\mathrm{T}} + \cdots + A_{0k} A_{0k}^{\mathrm{T}}, \tag{4.9}$$

由此

$$\mathrm{tr} R^{\mathrm{T}} R = \mathrm{tr} R R^{\mathrm{T}} + \mathrm{tr} A_{01} A_{01}^{\mathrm{T}} + \cdots + \mathrm{tr} A_{0k} A_{0k}^{\mathrm{T}} .$$

注意到

$$\mathrm{tr}R^{\mathrm{T}}R=\mathrm{tr}RR^{\mathrm{T}},$$

推出

$$\mathrm{tr}A_{01}A_{01}^{\mathrm{T}}+\cdots+\mathrm{tr}A_{0k}A_{0k}^{\mathrm{T}}=0 \tag{4.10}$$

显然,每个 $A_{0j}A_{0j}^{\mathrm{T}}$ 是半正定矩阵,必有 $\mathrm{tr}A_{0j}A_{0j}^{\mathrm{T}}\geq 0$,于是,从(4.10)有

$$\mathrm{tr}A_{0j}A_{0j}^{\mathrm{T}}=0,\quad j=1,\cdots,k.$$

再依 **3.6.7**,得

$$A_{0j}A_{0j}^{\mathrm{T}}=0,\quad j=1,\cdots,k.$$

因为 $A_{0j}A_{0j}^{\mathrm{T}}$ 的第 i 个主对角元素是 A_{0j} 的第 i 行中各(实)元素的平方和,所有这些主对角元素都必须为零,推出

$$A_{0j}=0,\quad j=1,\cdots,k, \tag{4.11}$$

并因而

$$R^{\mathrm{T}}R=RR^{\mathrm{T}},$$

因为 R 是上三角矩阵,在 **2.4.4** 的证明中已推出如此矩阵必为对角矩阵,因此

$$R=\mathrm{diag}(\lambda_1,\cdots,\lambda_p). \tag{4.12}$$

现在,考虑等式(4.9)两端的 2×2 主对角块.对于相应(4.7)中的块 A_{11},利用(4.11),得

$$A_{11}^{\mathrm{T}}A_{11}=A_{11}A_{11}^{\mathrm{T}}+A_{12}A_{12}^{\mathrm{T}}+\cdots+A_{1k}A_{1k}^{\mathrm{T}}. \tag{4.13}$$

注意到

$$\mathrm{tr}A_{11}^{\mathrm{T}}A_{11}=\mathrm{tr}A_{11}A_{11}^{\mathrm{T}},$$

推出

$$\mathrm{tr}A_{12}A_{12}^{\mathrm{T}}+\cdots+\mathrm{tr}A_{1k}A_{1k}^{\mathrm{T}}=0.$$

由此,因

$$\mathrm{tr}\left(A_{1j}A_{1j}^{\mathrm{T}}\right)\geq 0,\quad j=2,\cdots,k,$$

有

$$\mathrm{tr}\left(A_{1j}A_{1j}^{\mathrm{T}}\right)=0 \Rightarrow A_{1j}A_{1j}^{\mathrm{T}}=0 \Rightarrow A_{1j}=0,\ j=2,\cdots,k.$$

这样,(4.13)简化为

$$A_{11}^{\mathrm{T}}A_{11}=A_{11}A_{11}^{\mathrm{T}},$$

即 A_{11} 是正规矩阵.

如此,对 $j=2,\cdots,k-1$,依此考察等式(4.9)两端相应(4.7)中 A_{jj} 的 2×2 主对角块,利用同样的论证,推出(4.7)中所有非对角块是零,而所有对角块 A_{ii} 是正规矩阵.

至此,剩下要证明的是每个对角块 A_{jj} 形如(4.6).设

$$A_{jj}=\begin{bmatrix} a & b \\ c & d \end{bmatrix}\in\mathbb{R}^{2\times 2},$$

它是正规矩阵,于是从

$$A_{jj}^{\mathrm{T}}A_{jj}=A_{jj}A_{jj}^{\mathrm{T}},$$

得

$$b^2=c^2,\quad ac+bd=ab+cd. \tag{4.14}$$

注意,$c=b$ 情形和 $b=0$ 情形可以排除,因为这两种情形下 A_{jj} 是实对称的,只有实特征值;而根据(4.7)中的构造,每个块 A_{jj} 有一对共轭的非实特征值.

因此,$c=-b$,然后从(4.14)的第二个关系式即可推出 $a=d$.于是,每个块 A_{jj} 形如

$$\begin{bmatrix} a & b \\ -b & a \end{bmatrix},$$

即形如(4.6).

最后指出,这样的实矩阵有一对共轭的复特征值 $\lambda=a+\mathrm{i}b$ 和 $\bar{\lambda}=a-\mathrm{i}b$. □

2.4.9 定理 设 $A\in\mathbb{R}^{n\times n}$,则

(1) $A=A^{\mathrm{T}}$ 的充分必要条件是存在实正交矩阵 $Q\in\mathbb{R}^{n\times n}$,使

得

$$Q^{\mathrm{T}}AQ=\begin{pmatrix}\lambda_1 & 0 & \cdots & 0\\ 0 & \ddots & \ddots & \vdots\\ \vdots & \ddots & \ddots & 0\\ 0 & \cdots & 0 & \lambda_n\end{pmatrix},\quad \lambda_1,\cdots,\lambda_n\in\mathbb{R}。\tag{4.15}$$

(2) $A=-A^{\mathrm{T}}$ 的充分必要条件是存在实正交矩阵 $Q\in\mathbb{R}^{n\times n}$,使得

$$Q^{\mathrm{T}}AQ=\begin{bmatrix}0 & 0 & \cdots & 0\\ 0 & A_1 & \ddots & \vdots\\ \vdots & \ddots & \ddots & 0\\ 0 & \cdots & 0 & A_k\end{bmatrix},\tag{4.16}$$

其中每个 $A_j\in\mathbb{R}^{2\times 2}$ 形如

$$A_j=\begin{bmatrix}0 & \beta_j\\ -\beta_j & 0\end{bmatrix}.\tag{4.17}$$

(3) $AA^{\mathrm{T}}=I$ 的充分必要条件是存在实正交矩阵 $Q\in\mathbb{R}^{n\times n}$,使得

$$Q^{\mathrm{T}}AQ=\begin{bmatrix}\lambda_1 & o & \cdots & \cdots & \cdots & 0\\ 0 & \ddots & \ddots & & & \vdots\\ \vdots & \ddots & \lambda_p & \ddots & & \vdots\\ \vdots & & \ddots & A_1 & \ddots & \vdots\\ \vdots & & & \ddots & \ddots & 0\\ 0 & \cdots & \cdots & \cdots & 0 & A_k\end{bmatrix},\tag{4.18}$$

其中每个 $\lambda_i=\pm 1$,每个 $A_j\in\mathbb{R}^{2\times 2}$ 形如

$$A_j=\begin{bmatrix}\cos\theta_j & \sin\theta_j\\ -\sin\theta_j & \cos\theta_j\end{bmatrix},\quad \theta_j\in\mathbb{R}.\tag{4.19}$$

证 在每种情形中,所设条件已保证 A 是实正规矩阵,所以 A 可以表示成(4.5)和(4.6).

若 $A=A^{\mathrm{T}}$,则每个 A_j 成立

$$A_j=A_j^{\mathrm{T}},$$

故所有 β_j 均为零,$Q^{\mathrm{T}}AQ$ 是对角矩阵.

若 $A=-A^{\mathrm{T}}$,则每个 λ_j 成立

$$\lambda_j=-\lambda_j,$$

每个 A_j 成立

$$A_j=-A_j^{\mathrm{T}},$$

故所有 λ_j 均为零,所有 α_j 均为零.

若 $AA^{\mathrm{T}}=I$,则每个 λ_j 满足

$$\lambda_j\lambda_j=1,$$

从而 $\lambda_j=\pm1$;每个 A_j 满足

$$A_jA_j^{\mathrm{T}}=I,$$

从而 $\alpha_j^2+\beta_j^2=1$,可以写成 $\alpha_j=\cos\theta_j$ 和 $\beta_j=\sin\theta_j$. □

2.4.10 定义 设 $A\in\mathbb{C}^{n\times n}$,$\lambda\in\lambda(A)$ 称为矩阵 A 的**正规特征值**(normal eigenvalue),如果

(1) A 相应 λ 的各特征向量正交于 A 相应每个不同于 λ 的特征值的各特征向量.

(2) λ 的几何重数等于 λ 的代数重数.

2.4.11 定理 (1) 正规矩阵的每个特征值是正规特征值.

(2) 若 $A\in\mathbb{C}^{n\times n}$ 有 $n-1$ 个正规特征值(按重数计),则 A 是正规矩阵.

(3) λ 为 $A\in\mathbb{C}^{n\times n}$ 的正规特征值的充分必要条件是对每个酉矩阵 $U\in\mathbb{C}^{n\times n}$,$\lambda$ 为 U^*AU 的正规特征值.也就是说,正规特征值这

一性质是酉相似不变的.

证明留作练习.

2.5 条件数和病态矩阵

2.5.1 引言 粗略地讲,一个问题称为**良态**(well-condition)的,如果其参数的小的改变只会引起解的小的改变;一个问题称为**病态**(ill-condition)的,如果其参数的小的改变相反地会引起解的大的改变.这里扼要介绍一下矩阵求逆和线性方程组求解的误差估计问题.实际计算中会遇到种种误差.

设 $A\in\mathbb{C}^{n\times n}$ 是非奇异矩阵.

先考虑计算逆矩阵 A^{-1}.假设实际计算的是 $(A+\delta A)^{-1}$,其中 $\delta A\in\mathbb{C}^{n\times n}$ 足够"小",以致 $A+\delta A$ 是可逆的.因此,误差是

$$\delta A^{-1}\equiv(A+\delta A)^{-1}-A^{-1}=\left(I+A^{-1}\delta A\right)^{-1}A^{-1}-A^{-1}. \tag{5.1}$$

如果

$$\rho\left(A^{-1}\delta A\right)<1,$$

那么 $A+\delta A$ 可逆,而且可以将 $\left(I+A^{-1}\delta A\right)^{-1}$ 表示成关于 $A^{-1}\delta A$ 的幂级数.于是,有

$$\begin{aligned}\delta A^{-1}&=\sum_{k=0}^{\infty}(-1)^k\left(A^{-1}\delta A\right)^k A^{-1}-A^{-1}\\&=\sum_{k=1}^{\infty}(-1)^k\left(A^{-1}\delta A\right)^k A^{-1}.\end{aligned} \tag{5.2}$$

为了进行估计,任取一种矩阵的算子范数 $\|\bullet\|$.假设

$$\left\|A^{-1}\delta A\right\|<1,$$

这可以保证$\rho\left(A^{-1}\delta A\right)<1$和(5.2)成立.因而

$$\begin{aligned}\left\|\delta A^{-1}\right\|&=\left\|\sum_{k=1}^{\infty}(-1)^{k}\left(A^{-1}\delta A\right)^{k}A^{-1}\right\|\\&\le\sum_{k=1}^{\infty}\left\|A^{-1}\delta A\right\|^{k}\left\|A^{-1}\right\|=\frac{\left\|A^{-1}\delta A\right\|}{1-\left\|A^{-1}\delta A\right\|}\left\|A^{-1}\right\|.\end{aligned}$$

由此得到矩阵A求逆的相对误差的一个上界:

$$\frac{\left\|\delta A^{-1}\right\|}{\left\|A^{-1}\right\|}\le\frac{\left\|A^{-1}\delta A\right\|}{1-\left\|A^{-1}\delta A\right\|}. \tag{5.3}$$

如果再进一步假设

$$\|\delta A\|<1/\left\|A^{-1}\right\|,$$

则

$$\rho\left(A^{-1}\delta A\right)\le\left\|A^{-1}\delta A\right\|\le\left\|A^{-1}\right\|\|\delta A\|<1.$$

并从(5.3)有误差估计

$$\frac{\left\|\delta A^{-1}\right\|}{\left\|A^{-1}\right\|}\le\frac{\left\|A^{-1}\right\|\|\delta A\|}{1-\left\|A^{-1}\right\|\|\delta A\|}.$$

可将其改写成

$$\frac{\left\|\delta A^{-1}\right\|}{\left\|A^{-1}\right\|}\le\frac{\left\|A^{-1}\right\|\|A\|}{1-\left\|A^{-1}\right\|\|A\|\left(\|\delta A\|/\|A\|\right)}\frac{\|\delta A\|}{\|A\|}, \tag{5.4}$$

其中$\|\delta A\|/\|A\|$是A的(原始资料的)相对误差.

现在考虑求解线性方程组

$$Ax=b,\ b\in\mathbb{C}^{n}. \tag{5.5}$$

上面关于矩阵求逆的分析,同样可以应用来给出方程组实际所得解的精度的先验界.

假设实际求解的是方程组

$$(A+\delta A)(x+\delta x)=b+\delta b,\ \ \delta A\in\mathbb{C}^{n\times n},\ \ \delta x\in\mathbb{C}^{n},\quad(5.6)$$

其解 $x+\delta x$ 中, $x\in\mathbb{C}^n$ 是原始方程组(5.5)的解,而 $\delta x\in\mathbb{C}^n$ 是因 δA 和 δb 造成的误差.在条件

$$\rho\left(A^{-1}\delta A\right)<1$$

下,利用(5.2),有

$$\begin{aligned}\delta x&=(x+\delta x)-x=(A+\delta A)^{-1}(b+\delta b)-A^{-1}b\\&=\left[(A+\delta A)^{-1}-A^{-1}\right]b+(A+\delta A)^{-1}\delta b\\&=\sum_{k=1}^{\infty}(-1)^k\left(A^{-1}\delta A\right)^kA^{-1}b+\sum_{k=0}^{\infty}(-1)^k\left(A^{-1}\delta A\right)^kA^{-1}\delta b\\&=\sum_{k=1}^{\infty}(-1)^k\left(A^{-1}\delta A\right)^kx+\sum_{k=0}^{\infty}(-1)^k\left(A^{-1}\delta A\right)^kA^{-1}\delta b.\end{aligned}$$

因此,在条件

$$\left\|A^{-1}\delta A\right\|<1$$

下,

$$\begin{aligned}\|\delta x\|&\le\sum_{k=1}^{\infty}\left\|A^{-1}\delta A\right\|^k\|x\|+\sum_{k=0}^{\infty}\left\|A^{-1}\delta A\right\|^k\left\|A^{-1}\right\|\|\delta b\|\\&=\frac{\left\|A^{-1}\delta A\right\|}{1-\left\|A^{-1}\delta A\right\|}\|x\|+\frac{\left\|A^{-1}\right\|}{1-\left\|A^{-1}\delta A\right\|}\|\delta b\|.\end{aligned}$$

这里,关于向量和矩阵分别使用的是相容的向量范数和矩阵算子范数.注意到

$$\|b\|=\|Ax\|\le\|A\|\|x\|,$$

方程组(5.5)的解的相对误差为

$$\frac{\|\delta x\|}{\|x\|}\le\frac{\left\|A^{-1}\delta A\right\|}{1-\left\|A^{-1}\delta A\right\|}+\frac{\left\|A^{-1}\right\|\|A\|}{1-\left\|A^{-1}\delta A\right\|}\frac{\|\delta b\|}{\|A\|\|x\|}.$$

于是,在条件

$$\|A^{-1}\|\|\delta A\| < 1$$

下,有

$$\frac{\|\delta x\|}{\|x\|} \le \frac{\|A^{-1}\|\|A\|}{1-\|A^{-1}\|\|A\|(\|\delta A\|/\|A\|)}\left(\frac{\|\delta A\|}{\|A\|}+\frac{\|\delta b\|}{\|b\|}\right), \tag{5.7}$$

其中 $\|\delta A\|/\|A\|$ 仍为 A 的相对误差,而 $\|\delta b\|/\|b\|$ 是 b 的相对误差.

从(5.4)和(5.7)看出,当 $\|\delta A\|$ 和 $\|\delta b\|$ 足够"小"时,(5.4)的右端的阶为

$$\|A^{-1}\|\|A\|\frac{\|\delta A\|}{\|A\|},$$

(5.7)的右端的阶为

$$\|A^{-1}\|\|A\|\left(\frac{\|\delta A\|}{\|A\|}+\frac{\|\delta b\|}{\|b\|}\right).$$

以上分析足见数

$$\|A^{-1}\|\|A\|$$

在计算 A^{-1} 和求解方程组 $Ax = b$ 的误差分析中起着关键的作用.

2.5.2 定义 设 $A \in \mathbb{C}^{n\times n}$,$\|\bullet\|$ 是一种算子范数,

$$\kappa(A) \equiv \begin{cases} \|A^{-1}\|\|A\|, & \det A \ne 0, \\ \infty, & \det A = 0, \end{cases} \tag{5.8}$$

称 $\kappa(A)$ 为矩阵 A 求逆的关于算子范数 $\|\bullet\|$ 的**条件数**(condition number).

常用的是关于 p－范数 $\|\bullet\|_p$ 的条件数,可记作 $\kappa_p(A)$.特别,称 $\kappa_2(A)$ 为**谱条件数**.

利用条件数,矩阵求逆的误差估计(5.4)可表示成

$$\frac{\|\delta A^{-1}\|}{\|A^{-1}\|}\le\frac{\kappa(A)}{1-\kappa(A)\frac{\|\delta A\|}{\|A\|}}\frac{\|\delta A\|}{\|A\|},\tag{5.9}$$

方程组解的误差估计(5.7)可表示成

$$\frac{\|\delta x\|}{\|x\|}\le\frac{\kappa(A)}{1-\kappa(A)\frac{\|\delta A\|}{\|A\|}}\left(\frac{\|\delta A\|}{\|A\|}+\frac{\|\delta b\|}{\|b\|}\right)\tag{5.10}$$

2.5.3 定理 设 $A\in\mathbb{C}^{n\times n}$ 是非奇异矩阵，则

(1) 成立

$$\kappa(A)\ge 1;$$
$$\kappa(A)=\kappa(A^{-1});$$
$$\kappa(A)=\kappa(\alpha A),\quad \alpha\in\mathbb{C},\ \alpha\ne 0.$$

(2) $\kappa_2(A)=\dfrac{\sigma_1(A)}{\sigma_n(A)}$,其中 $\sigma_1(A)$ 和 $\sigma_n(A)$ 分别表示 A 的最大奇异值和最小奇异值,即 A^*A 的最大特征值和最小特征值的非负平方根.

(3) 若 A 是正规矩阵,则

$$\kappa_2(A)=\frac{\max\limits_{\lambda\in\lambda(A)}|\lambda|}{\min\limits_{\lambda\in\lambda(A)}|\lambda|}.$$

(4) 若 A 是酉矩阵,则 $\kappa_2(A)=1$.

(5) $\kappa_2(A)$ 是酉不变的,即对任何酉矩阵 $U\in\mathbb{C}^{n\times n}$ 成立

$$\kappa_2(A)=\kappa_2(UA)=\kappa_2(AU).$$

证 (1) 对任何一种矩阵算子范数 $\|\bullet\|$ 及任何 $\alpha\in\mathbb{C},\alpha\ne 0$,

$$\kappa(A)=\|A^{-1}\|\|A\|\ge\|A^{-1}A\|=\|I\|=1;$$

$$\kappa\left(A^{-1}\right)=\left\|\left(A^{-1}\right)^{-1}\right\|\left\|A^{-1}\right\|=\|A\|\left\|A^{-1}\right\|=\kappa(A);$$

$$\kappa(\alpha A)=\left\|(\alpha A)^{-1}\right\|\|\alpha A\|$$

$$=|\alpha|^{-1}\left\|A^{-1}\right\||\alpha|\|A\|=\kappa(A).$$

(2) 直接从$\kappa_2(A)=\left\|A^{-1}\right\|_2\|A\|_2$及 **1.8.31** 推出.

(3) 直接从(2)及 **2.4.6** 推出.

(4) 直接从(3)及 **2.3.4** 推出.

(5) 直接从(2)及奇异值酉不变推出.奇异值酉不变是因为有

$$(UA)^*(UA)=A^*U^*UA=A^*A.$$ □

从(5.9)和(5.10)看出:如果$\|\delta A\|$和$\|\delta b\|$是小的,条件数$\kappa(A)$不大,那么矩阵求逆的相对误差$\left\|\delta A^{-1}\right\|/\left\|A^{-1}\right\|$和矩阵的相对误差$\|\delta A\|/\|A\|$是同阶的,方程组求解的相对误差$\|\delta x\|/\|x\|$与矩阵的相对误差及常数项向量相对误差之和$\left(\|\delta A\|/\|A\|\right)+\left(\|\delta b\|/\|b\|\right)$是同阶的.然而,如果$\kappa(A)$是大的,那么这就是一个不很好的信息.因此,有理由引入下面的定义.

2.5.4 定义 设$A\in\mathbb{C}^{n\times n}$,$\|\bullet\|$是一种矩阵范数,$\kappa(\bullet)$是关于$\|\bullet\|$的条件数.

如果$\kappa(A)$是大的,则说A是关于范数$\|\bullet\|$的矩阵求逆或方程组(5.5)求解的**病态矩阵**.

如果$\kappa(A)$是小的,则说A是关于范数$\|\bullet\|$的矩阵求逆或方程组(5.5)求解的**良态矩阵**;特别,$\kappa(A)=1$时,说A是**完全良态矩阵**.

2.5.5 定义 矩阵$H_n=[h_{ij}]\in\mathbb{R}^{n\times n}$,

$$h_{ij}=\frac{1}{i+j-1},\quad i,j=1,\cdots,n.$$

更具体地,

$$H_n \equiv \begin{bmatrix} 1 & \frac{1}{2} & \frac{1}{3} & \cdots & \frac{1}{n} \\ \frac{1}{2} & \frac{1}{3} & \frac{1}{4} & \cdots & \frac{1}{n+1} \\ \frac{1}{3} & \frac{1}{4} & \frac{1}{5} & \cdots & \frac{1}{n+2} \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ \frac{1}{n} & \frac{1}{n+1} & \frac{1}{n+2} & \cdots & \frac{1}{2n-1} \end{bmatrix}, \qquad (5.11)$$

H_n 称为 **Hilbert 矩阵**.

不难找到或举出病态矩阵的各种例子.然而,或许最常被引用的一个典型例子是 Hilbert 矩阵.

H_n 是对称的.由于

$$x^{\mathrm{T}} H_n x = \sum_{i,j=1}^{n} h_{ij} x_i x_j$$

$$= \sum_{i,j=1}^{n} \int_0^1 t^{i+j} x_i x_j \mathrm{d}t = \int_0^1 \left(\sum_{i=1}^{n} x_i t^i \right)^2 \mathrm{d}t > 0,$$

$$\forall x = (x_1, \cdots, x_n)^{\mathrm{T}} \in \mathbb{R}^n, x \neq 0, \qquad (5.12)$$

因而 H_n 还是正定的.而且,H_n 的每个元素 $h_{ij} \in (0,1]$;谱半径也不大,当 $n \to \infty$ 时,

$$\rho(H_n) = \pi + O(1/\log n). \qquad (5.13)$$

但是,H_n 却是非常病态的.依 **2.5.3** 的(3),

$$\kappa_2(H_n) = \lambda_{\max}(H_n)/\lambda_{\min}(H_n).$$

可以证明,随着 n 的增大,

$$\kappa_2(H_n) \approx \mathrm{e}^{\alpha n}, \quad \alpha \approx 3.5. \qquad (5.14)$$

例如,

$$\kappa_2(H_3) \approx 5\times 10^2,$$
$$\kappa_2(H_6) \approx 1.5\times 10^7,$$
$$\kappa_2(H_8) \approx 1.5\times 10^{10}.$$

还可以证明: $H_n^{-1} = \left[\hat{h}_{ij}\right]$,

$$\hat{h}_{ij} = \frac{(-1)^{i+j}(n+i-1)!(n+j-1)!}{[(i-1)!(j-1)!]^2(n-i)!(n-j)!(i+j-1)}, \quad i,j=1,\cdots,n. \tag{5.15}$$

2.5.6 定理 设 $A, \delta A \in \mathbb{C}^{n\times n}$, A 是可对角化的,即存在非奇异矩阵 $S \in \mathbb{C}^{n\times n}$,使得

$$A = S\Lambda S^{-1}, \quad \Lambda = \mathrm{diag}(\lambda_1,\cdots,\lambda_n).$$

并设 $\|\bullet\|$ 是一种矩阵范数,满足

$$\|D\| = \max_{1\le i\le n}|d_i|, \quad \forall D = \mathrm{diag}(d_1,\cdots,d_n),$$

$\hat{\lambda} \in \mathbb{C}$ 是 $A+\delta A$ 的一个特征值,则有 A 的某个特征值 λ_i,成立

$$\|\hat{\lambda} - \lambda_i\| \le \|S\|\|S^{-1}\|\|\delta A\| = \kappa(S)\|\delta A\|. \tag{5.16}$$

证 考虑

$$S^{-1}(A+\delta A)S = \Lambda + S^{-1}(\delta A)S$$

的特征值.因 $\hat{\lambda}$ 是 $\Lambda + S^{-1}(\delta A)S$ 的特征值,故

$$\hat{\lambda}I - \Lambda - S^{-1}(\delta A)S$$

是奇异的.

如果 $\hat{\lambda}I - \Lambda$ 是奇异的,那么对某个 i, $\hat{\lambda} = \lambda_i$,自然满足 (5.16).

然而,假若 $\hat{\lambda}I - \Lambda$ 是非奇异的.此时,矩阵

$$\left(\hat{\lambda}-\Lambda\right)^{-1}\left(\hat{\lambda}I-\Lambda-S^{-1}(\delta A)S\right)=I-\left(\hat{\lambda}I-\Lambda\right)^{-1}S^{-1}(\delta A)S$$

是奇异的.依 **1.8.27**,必有

$$\left\|\left(\hat{\lambda}I-\Lambda\right)^{-1}S^{-1}(\delta A)S\right\|\geq 1.$$

因此,基于矩阵范数$\|\bullet\|$关于对角矩阵的性态的假设,有

$$1\leq\left\|\left(\hat{\lambda}I-\Lambda\right)^{-1}S^{-1}(\delta A)S\right\|\leq\left\|S^{-1}(\delta A)S\right\|\left\|\left(\hat{\lambda}I-\Lambda\right)^{-1}\right\|$$

$$=\left\|S^{-1}(\delta A)S\right\|\max_{1\leq i\leq n}\frac{1}{\left|\hat{\lambda}-\lambda_i\right|}=\frac{\left\|S^{-1}(\delta A)S\right\|}{\min\limits_{1\leq i\leq n}\left|\hat{\lambda}-\lambda_i\right|},$$

由此,得

$$\min_{1\leq i\leq n}\left|\hat{\lambda}-\lambda_i\right|\leq\left\|S^{-1}(\delta A)S\right\|\leq\left\|S^{-1}\right\|\left\|\delta A\right\|\left\|S\right\|=\kappa(S)\left\|\delta A\right\|. \quad \square$$

2.6 Vandermonde 矩阵及 Cauchy 矩阵

2.6.1 定义 设$x_1,\cdots,x_n\in\mathbb{C}$,矩阵

$$V(x_1,\cdots,x_n)\equiv\begin{bmatrix}1&\cdots&1&\cdots&1\\x_1&\cdots&x_j&\cdots&x_n\\x_1^2&\cdots&x_j^2&\cdots&x_n^2\\\vdots&\vdots&\vdots&\vdots&\vdots\\x_1^{n-1}&\cdots&x_j^{n-1}&\cdots&x_n^{n-1}\end{bmatrix}\in\mathbb{C}^{n\times n}, \quad (6.1)$$

其中第j列是比值为x_j的等比序列,$j=1,\cdots,n$.$V(x_1,\cdots,x_n)$称为关于数组$x_1,\cdots,x_n$的 **Vandermonde 矩阵**.

Vandermonde 矩阵的最通常的应用例子是它出现在插值多项

式的存在性和唯一性的问题中.多项式插值问题的最基本的提法是:设有 n 个数对

$$\left(x_i, f_i\right), \quad i=1,\cdots,n,$$

$x_1,\cdots,x_n$ 是不同的数;求一个次数小于等于 n 的多项式

$$p(x)=a_0+a_1x+\cdots+a_{n-1}x^{n-1},$$

使得

$$p\left(x_i\right)\equiv a_0+a_1x_i+\cdots+a_{n-1}x_i^{n-1}=f_i, \quad i=1,\cdots,n.$$

这是以 $a_0,\cdots,a_{n-1}$ 为未知数的线性方程组,其系数矩阵,即(6.1)所示的 $V\left(x_1,\cdots,x_n\right)$ 的转置.

2.6.2 定理 Vandermonde 矩阵的行列式

$$\det V\left(x_1,\cdots,x_n\right)=\prod_{j>i}\left(x_j-x_i\right)=\prod_{i=1}^{n}\prod_{j=i+1}^{n}\left(x_j-x_i\right). \quad (6.2)$$

证 利用 **1.5.6** 的行列式的公式,得知 $\det V\left(x_1,\cdots,x_n\right)$ 是关于 $x_1,\cdots,x_n$ 的次数小于等于 $n(n-1)/2$ 的 n 元多项式.设想无论哪两个数 x_i 和 x_j 相等,$i\neq j$,那么 $V\left(x_1,\cdots,x_n\right)$ 就有两列相等,其行列式等于零.所以,根据代数学因式分解定理,$\det V\left(x_1,\cdots,x_n\right)$ 可被 x_j-x_i 整除.由此推出,$\det V\left(x_1,\cdots,x_n\right)$ 可被积

$$\prod_{j>i}\left(x_j-x_i\right)$$

整除.此积的次数为 $n(n-1)/2$.因此

$$\det V\left(x_1,\cdots,x_n\right)=c_n\prod_{j>i}\left(x_j-x_i\right), \quad (6.3)$$

c_n 是一个常数.依 **1.5.11**,对 $\det V\left(x_1,\cdots,x_n\right)$ 关于最后一列即 $j=n$ 作 Laplace 展开,得到按 x_n 的幂的展开式,x_n^{n-1} 的系数是

$$\det V\left(x_1,\cdots,x_{n-1}\right).$$

另一方面,在等式(6.3)的右端,x_n^{n-1} 的系数是

$$c_n \prod_{n>j>i}(x_j - x_i).$$

这样,对 $V(x_1,\cdots,x_{n-1})$ 应用(6.3),得

$$\det V(x_1,\cdots,x_{n-1}) = c_{n-1}\prod_{j>i}(x_j - x_i) = c_n \prod_{n>j>i}(x_j - x_i),$$

推出

$$c_n = c_{n-1},$$

显然,$c_2 = 1$;因此,用数学归纳法得到对所有 n 有 $c_n = 1$,故(6.2)成立. □

2.6.3 定义 设矩阵

$$F \equiv V(1,\omega,\omega^2,\cdots,\omega^{n-1})$$

$$= \begin{bmatrix} 1 & 1 & 1 & \cdots & 1 \\ 1 & \omega & \omega^2 & \cdots & \omega^{n-1} \\ 1 & \omega^2 & \omega^4 & \cdots & \omega^{2(n-1)} \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ 1 & \omega^{n-1} & \omega^{2(n-1)} & \cdots & \omega^{(n-1)^2} \end{bmatrix} \in \mathbb{C}^{n\times n}, \quad (6.4)$$

其中

$$\omega = \mathrm{e}^{-\mathrm{j}\frac{2\pi}{n}}, \quad (6.5)$$

F 称为 **Fourier 矩阵**.

F 出现在离散 Fourier 分析和快速 Fourier 变换中.

F 是 Vandermonde 矩阵的特殊情形,可表示为

$$F = [f_{rs}] \in \mathbb{C}^{n\times n}, \quad (6.6)$$

其中

$$f_{rs} = \omega^{(r-1)(s-1)}, \quad r,s = 1,\cdots,n.$$

2.6.4 定理 设 $F \in \mathbb{C}^{n\times n}$ 是 Fourier 矩阵，则

(1) F 是对称矩阵.

(2) F 中只有 $1,\omega,\omega^2,\cdots,\omega^{n-1}$ 这 n 个不同元素.

(3) F 是酉矩阵.

证 (1) 显然.

(2) 因为 $\omega^n = 1$，$\{1,\omega,\omega^2,\cdots\}$ 是周期序列.

(3) 利用 $\omega\overline{\omega} = 1$，$\overline{\omega}^p = \omega^{-p}$，以及

$$1+\omega+\omega^2+\cdots+\omega^{n-1}=0, \tag{6.7}$$

即可推出 $F^*F = FF^* = I$. □

2.6.5 定义 设 $x_1,\cdots,x_s \in \mathbb{C}$; $k_1,\cdots,k_s$ 是正整数，而且满足 $k_1+\cdots+k_s=n$. 矩阵

$$\tilde{V}=\begin{bmatrix}W_1 & \cdots & W_s\end{bmatrix}\in\mathbb{C}^{n\times n},$$

其中

$$W_i=\begin{bmatrix}X_i^{(0)} & X_i^{(1)} & \cdots & X_i^{(k_i-1)}\end{bmatrix}\in\mathbb{C}^{n\times k_i},\quad i=1,\cdots,s.$$

而且

$$X_i^{(m)}=\frac{1}{m!}\frac{\mathrm{d}^m}{\mathrm{d}x_i^m}\left(1,x_i,x_i^2,\cdots,x_i^{n-1}\right)^{\mathrm{T}},$$

$$m=0,1,\cdots,k_i-1,\quad i=1,\cdots,s.$$

$\tilde{V}_n$ 称为**广义 Vandermonde 矩阵**.

不难看出，当 $s=n$ 时，$k_1=\cdots=k_n=1$，$\tilde{V}_n$ 就是形如(6.1)的 Vandermonde 矩阵.

2.6.6 定义 设 $a_1,\cdots,a_n \in \mathbb{C}$ 和 $b_1,\cdots,b_n \in \mathbb{C}$，矩阵

$$C(a,b)\equiv C(a_1,\cdots,a_n;b_1,\cdots,b_n)$$

$$
= \begin{bmatrix} \frac{1}{a_1+b_1} & \frac{1}{a_1+b_2} & \cdots & \frac{1}{a_1+b_n} \\ \frac{1}{a_2+b_1} & \frac{1}{a_2+b_2} & \cdots & \frac{1}{a_2+b_n} \\ \vdots & \vdots & \vdots & \vdots \\ \frac{1}{a_n+b_1} & \frac{1}{a_n+b_2} & \cdots & \frac{1}{a_n+b_n} \end{bmatrix}, \qquad (6.8)
$$

$C(a,b)$称为关于数组$a_1,\cdots,a_n$和数组$b_1,\cdots,b_n$的 **Cauchy 矩阵**.

2.6.7 定理 Cauchy 矩阵的行列式为

$$
\det C(a,b) = \frac{\prod_{j>i}(a_j-a_i)(b_j-b_i)}{\prod_{i,j=1}^{n}(a_i+b_j)}. \qquad (6.9)
$$

证 对于$C(a,b)$的行列式,利用 **1.5.6** 的公式,并且对其各项采用公分母,可以写成

$$
\det C(a,b) = \frac{p(a,b)}{\prod_{i,j=1}^{n}(a_i+b_j)}, \qquad (6.10)
$$

其中p是关于$a_1,\cdots,a_n$和$b_1,\cdots,b_n$的次数小于等于n^2-n的多项式.

设想当$a_i=a_j$时,$C(a,b)$的第i行和第j行相等.另外,当$b_i=b_j$时,$C(a,b)$的第i列和第j列相等.无论哪种情形,都有$\det C(a,b)=0$;所以,根据代数学的因式分解定理,多项式p可被a_j-a_i和b_j-b_i整除,并因此可被积

$$
\prod_{j>i}(a_j-a_i)(b_j-b_i)
$$

整除.此积的次数正好和 p 的次数相同,为 n^2-n;于是

$$p(a,b)=c_n\prod_{j>i}(a_j-a_i)(b_j-b_i), \tag{6.11}$$

其中 c_n 是常数.依 **1.5.11**,对 $C(a,b)$ 关于最后一列即 $j=n$ 作 Laplace 展开,相应于 $1/(a_n+b_n)$ 的项是

$$\det C(a_1,\cdots,a_{n-1};b_1,\cdots,b_{n-1})\frac{1}{(a_n+b_n)}.$$

现令 $a_n=b_n=d$,从(6.10)和(6.11)得

$$\det C(a_1,\cdots,a_{n-1},d;b_1,\cdots,b_{n-1},d)$$

$$=\frac{c_n\prod\limits_{n>i}(d-a_i)(d-b_i)}{2d\prod\limits_{n>i}(d+a_i)(d+b_i)}\frac{\prod\limits_{n>j>i}(a_j-a_i)(b_j-b_i)}{\prod\limits_{i,j<n}(a_i+b_j)} \tag{6.12}$$

另一方面,从 Laplace 展开式得

$$\det C(a_1,\cdots,a_{n-1},d;b_1,\cdots,b_{n-1},d)$$

$$=\frac{1}{2d}\det C(a_1,\cdots,a_{n-1};b_1,\cdots,b_{n-1})+\cdots(\text{其它项}) \tag{6.13}$$

利用等式(6.12)和(6.13)右端相等并乘以 $2d$,然后令 $d=0$;再对矩阵 $C(a_1,\cdots,a_{n-1};b_1,\cdots,b_{n-1})$ 应用(6.10)和(6.11),推出 $c_n=c_{n-1}$.显然,$c_1=1$.于是从数学归纳法推出对所有 n 有 $c_n=1$;这样,从(6.10)和(6.11)即得(6.9). □

3 自伴矩阵和稳定矩阵

3.1 二次型

3.1.1引言 设 f 是 n 个实变量 $x_1,\cdots,x_n$ 的二次连续可微实值函数,令

$$x=\left(x_1,\cdots,x_n\right)^{\mathrm{T}}, f(x)\equiv f\left(x_1,\cdots,x_n\right).$$

考虑 f 在点 $a=\left(a_1,\cdots,a_n\right)^{\mathrm{T}}$ 使用 Taylor 逼近至二阶:

$$f(a+y)=f(a)+l(y)+\frac{1}{2}q(y)+\|y\|^2\varepsilon(\|y\|), \quad (1.1)$$

其中 $y=\left(y_1,\cdots,y_n\right)^{\mathrm{T}}$;$\varepsilon$ 是某一函数:当 $t\to 0$ 时 $\varepsilon(t)\to 0$;l 是一线性函数,利用 Euclid 内积可以写成

$$l(y)=(y,g), \quad (1.2)$$

$g\equiv\left(g_1,\cdots,g_n\right)^{\mathrm{T}}$ 是 f 在点 a 的梯度,依照 Taylor 逼近的定理,

$$g_i=\left.\frac{\partial f}{\partial x_i}\right|_{x=a}, \quad i=1,\cdots,n. \quad (1.3)$$

q 是二次函数或称**二次型**,

$$q(y)=\sum_{i,j=1}^{n}h_{ij}y_iy_j. \quad (1.4)$$

矩阵 $H\equiv[h_{ij}]$ 称为 f 的 **Hesse 矩阵**,依照 Taylor 逼近的定理,

$$h_{ij}=\left.\frac{\partial^2 f}{\partial x_i\partial x_j}\right|_{x=a}, \quad i,j=1,\cdots,n. \quad (1.5)$$

利用矩阵记号和 Euclid 内积,(1.4)可以改写成

$$q(y)=(y,Hy). \tag{1.6}$$

由假定及(1.5)得知 h_{ij} 是实数,且

$$h_{ij}=h_{ji}, \quad i,j=1,\cdots,n. \tag{1.7}$$

这样,H 的伴随矩阵就是它自己,$H^*=H$,因此自然采用"自伴"一词,称 H 为**自伴**的.

进一步说,这里 H 是实矩阵,成立

$$H^*=\overline{H^{\mathrm{T}}}=H^{\mathrm{T}}$$

于是

$$H=H^{\mathrm{T}}. \tag{1.8}$$

一般地说,H 是实或复矩阵,满足(1.8),称其为**对称矩阵.**

现在假设 a 是 f 的临界点,则 $g=\operatorname{grad}f$ 在点 a 处为零;此时 Taylor 公式(1.1)表明,f 在点 a 附近的性态是由二次项 q 来控制的. 二次函数在临界点附近的性态对于动力系统以及在几何学中是有基本重要性的性态;这就是数学中会给予二次函数(二次型)以重要地位,以及使得关于对称矩阵的分析成为线性代数中重要课题的原因.

为了研究二次型 q,常采用引进新的变量 $z=(z_1,\cdots,z_n)^{\mathrm{T}}$:

$$Ly=z. \tag{1.9}$$

L 是某个非奇异矩阵,使得 q 按照 z 有一个简单的形式.

3.1.2 定理 对于给定的二次型(1.4):

(1) 存在形如(1.9)的变量替换使得 q 对角化,即形如

$$q(L^{-1}z)=\sum_{i=1}^{n}d_iz_i^2. \tag{1.10}$$

(2) 存在许多引进新变量对角化 q 的途径,并且不同途径产生

的如(1.10)中 $d_1,\cdots,d_n$ 那样的对角项系数一般是不同的,但是它们之中出现的正数、负数和零的个数全都相同.

证 (1)的证明全然是初等的和构造性的.先讨论三种情形:

第一种情形. 假设(1.4)的对角项之一不为零,比如说

$$h_{11} \neq 0,$$

那么将包含 y_1 的所有项集合在一起:

$$h_{11}y_1^2 + \sum_{j=2}^{n} h_{1j}y_1y_j + \sum_{i=1}^{n} h_{i1}y_iy_1 ,$$

利用 H 的对称性及配平方,可将上式写成

$$h_{11}\left(y_1 + h_{11}^{-1}\sum_{j=2}^{n} h_{1j}y_j\right)^2 - h_{11}^{-1}\left(\sum_{j=2}^{n} h_{1j}y_j\right)^2 .$$

令

$$y_1 + h_{11}^{-1}\sum_{j=2}^{n} h_{1j}y_j = z_1, \tag{1.11}$$

于是,q 可以写成如下形式:

$$q(y) = h_{11}z_1^2 + q_2(y), \tag{1.12}$$

其中 q_2 仅依赖于 $y_2,\cdots,y_n$.

第二种情形. 假设(1.4)的对角项全为零但非对角项之一不为零,比如说

$$h_{12} = h_{21} \neq 0,$$

那么将 y_1+y_2 和 y_1-y_2 作为新变量便可产生非零对角项,从而归结为第一种情形.

第三种情形. 假设(1.4)的所有对角项和非对角项全为零,那么 $q(y)\equiv 0$,也就没有什么可证的了.

这样,剩下的证明就很简单了,只须对变量个数 n 应用数学归纳法.然而(1.12)已经表明,如果作归纳法假设“$n-1$ 个变量的二次型能写成(1.10)的形式”,那么 q 本身就能写成(1.10)的形式.而且,从 $y_2,\cdots,y_n$ 通过一个可逆矩阵与 $z_2,\cdots,z_n$ 相联系,以及(1.11),立即推

出向量 y 能通过一个可逆矩阵与向量 z 相联系.

现在证明(2).考虑 q 在坐标 $z_1,\cdots,z_n$ 下的表示式(1.10),引进记号

$$i_+, i_-, i_0$$

分别表示系数 $d_1,\cdots,d_n$ 中正、负、零的个数.假设 $z_1,\cdots,z_n$ 经适当标号使得 $d_1,\cdots,d_n$ 中前 $i=i_+$ 个 $d_1,\cdots,d_i$ 是正数,而其余的是非正数.

设 $S\subset\mathbb{R}^n$ 是子空间,如果成立

$$q(u)>0,\quad \forall u\in S, u\neq 0, \tag{1.13}$$

则称 q 在子空间 S 上是**正**的.下面证明

$$i_+=\max\{\dim S: S\subset\mathbb{R}^n \text{是子空间且} q \text{在} S \text{上为正}\}. \tag{1.14}$$

定义子空间 $S_+\subset\mathbb{R}^n$,

$$S_+\equiv\left\{z=(z_1,\cdots,z_n)^{\mathrm{T}}: z_{i+1}=\cdots=z_n=0\right\},$$

显然,$\dim S_+=i_+$,而且 q 在子空间 S_+ 上是正的.因此 i_+ 小于等于(1.14)的右边.另一方面,假设有子空间 $S\subset\mathbb{R}^n$ 使 q 在 S 上是正的且 $\dim S>i_+$;作映射 P,

$$Pz\equiv(z_1,\cdots,z_i,0\cdots,0),\quad \forall z=(z_1,\cdots,z_n)^{\mathrm{T}}\in S,$$

P 的目标空间的维数是 i_+,它小于域空间的维数.依 **1.3.4**,在 P 的零空间中存在非零向量 y.依 P 的定义,y 向量在坐标 $z_1,\cdots,z_n$ 下的前 i_+ 个分量为零.于是从(1.10)推出 $q(y)\le 0$;这与 q 在 S 上为正的假设矛盾,所以(1.14)成立.

类似地,可以定义所谓"q 在子空间 S 上是**负的**",而且相仿可以证明

$$i_-=\max\{\dim S: S\subset\mathbb{R}^n \text{是子空间且} q \text{在} S \text{上为负}\}. \tag{1.15}$$

至此,证明了 i_+ 和 i_- 可以通过二次型 q 本身来定义,本质上与把

q 转变成(1.10)形式的变量的选取无关.这样,$i_0 = n - i_+ - i_-$ 自然也与变量的选取无关. □

该定理的(2)称为 **Sylvester 惯性律**.

利用 q 的表达形式(1.6),可以用矩阵术语重新解释上述定理.用 $M \equiv L^{-1}$ 左乘(1.9),得

$$y = Mz, \tag{1.16}$$

将(1.16)代入(1.6),得

$$q(y) = (y, Hy) = (Mz, HMz) = (z, M^*HMz), \tag{1.17}$$

显然,q 对于 z 形如(1.10)当且仅当 M^*HM 为对角矩阵.因此上述定理的(1)可以转述成下面的定理.

3.1.3 定理 设 H 是实自伴矩阵,则存在可逆矩阵 M 使得

$$M^*HM = D \tag{1.18}$$

其中 D 是对角矩阵. □

在许多应用中,对于形如(1.16)的变换变量来说,极其重要的是在旧的和新的变量下能保持 Euclid 长度不变:$\|y\| = \|z\|$.依 **1.7.25** 和 **1.7.26**,这意味着 M 是正交矩阵,即满足

$$M^*M = I. \tag{1.19}$$

不仅属于线性代数而且还属于数学总体的基本定理之一是:每个实值二次型 q 可以通过变量的等距变换使其对角化.换用矩阵语言来说是:每个实对称矩阵 H,存在实可逆矩阵 M,使得(1.18)和(1.19)同时成立.

后面将给出这一重要结果的两种证明.第一种证明基于在 **1.6** 提供的一般矩阵的谱论,将其应用于复 Euclid 空间的自伴映射. 第二种证明见 **3.5.2**.

3.2 自伴矩阵的基本性质和谱定理

3.2.1 定义 设 X 是复 Euclid 空间,$H \in \mathscr{L}(X,X)$,即 H 是 X 到 X 的线性映射,H^* 是 H 的伴随映射.如果满足

$$H^* = H, \tag{2.1}$$

则称 H 为**自伴映射**.

平行地,对于矩阵来说,满足形如(2.1)关系式的 $H \in \mathbb{C}^{n\times n}$ 称为**自伴矩阵**(selfadjoint matrix),常称为 **Hermite 矩阵**.

以下集中论述自伴矩阵,它们多可换用映射语言.

现设 $H \in \mathbb{C}^{n\times n}$ 是自伴矩阵,则依 **1.7.19** 及(2.1),成立

$$(Hx,y) = (x,H^*y) = (x,Hy), \quad \forall x,y \in \mathbb{C}^n. \tag{2.2}$$

关于自伴矩阵,有以下明显的性质(留作练习):

(1) 对于任意的矩阵 $A \in \mathbb{C}^{n\times n}$,$A + A^*$,$AA^*$ 和 A^*A 都是自伴矩阵.

(2) 若 $H \in \mathbb{C}^{n\times n}$ 是自伴矩阵,则 $H^k (k = 1,2,\cdots)$ 是自伴矩阵.

(3) 若 $H \in \mathbb{C}^{n\times n}$ 是非奇异自伴矩阵,则 H^{-1} 也是自伴矩阵.

(4) 若 $H,G \in \mathbb{C}^{n\times n}$ 是自伴矩阵,$\alpha,\beta \in \mathbb{R}$,则 $\alpha H + \beta G$ 是自伴矩阵.

(5) 若 $H = [h_{ij}] \in \mathbb{C}^{n\times n}$ 是自伴矩阵,则

$$h_{ii} \in \mathbb{R}, \quad i = 1,\cdots,n.$$

(6) 若 H 是自伴矩阵,则 H 是正规矩阵.

3.2.2 定义 矩阵 $H \in \mathbb{C}^{n\times n}$ 称为**对称矩阵**,如果满足

$$H^{\mathrm{T}} = H\ ,\qquad (2.3)$$

其中H^{T}是H的转置矩阵.

显然,实对称矩阵一定是自伴矩阵,复对称矩阵则不然.

3.2.3 自伴矩阵/自伴映射谱定理 设$H \in \mathbb{C}^{n\times n}$是自伴矩阵,则$H$的特征值是实数;而且存在构成$\mathbb{C}^n$的正交基的特征向量组,或者说$H$是酉可对角化的,即存在酉矩阵$U \in \mathbb{C}^{n\times n}$,使得

$$H = UDU^*\ ,$$

其中D是以H的n个(按重数计)特征值为对角元素的对角矩阵.

换一种说法:设H是复 Euclid 空间X到X的自伴映射,则H的特征值是实数,而且存在构成X的正交基的特征向量组.

证 依谱论主要结果 **1.6.13**,H的特征向量和广义特征向量生成空间$\mathbb{C}^n$.为了从 **1.6.13** 推出本定理,需要证明自伴矩阵H必有如下性质:

(1) H只有实特征值.

(2) H没有广义特征向量,只有真特征向量.

(3) H的相应不同特征值的特征向量是正交的.

下面证明这些性质.

(1) 如果复数$\lambda \equiv a + \mathrm{i}b$是$H$的特征值,那么$\mathrm{i}b$是亦为自伴矩阵的$H - aI$的特征值.所以,只须证明自伴矩阵$H$没有纯虚数的特征值$\mathrm{i}b$.假定$\mathrm{i}b$是$H$的特征值,相应的特征向量为$z$:

$$Hz = \mathrm{i}bz\ ,$$

则一方面依(2.2),有

$$\left(Hz, z\right) = \left(z, Hz\right) = \overline{\left(Hz, z\right)},\qquad (2.4)$$

得知$\left(Hz, z\right)$是实数;另一方面,$\left(z, z\right)$也是实数,而且

$$\left(Hz, z\right) = \left(ibz, z\right) = ib\left(z, z\right),$$

因此只能$b = 0$.这就是说,H不可能有纯虚数的特征值.

(2) 假定H的特征值λ有广义特征向量z,那么$H - \lambda I$以0为特征值且有广义特征向量z.因此不妨假定H以0为特征值且有

广义特征向量 z,并满足

$$H^m z = 0.$$

下面证明必有 $Hz = 0$,即 z 是真特征向量.

先考虑 $m = 2$ 的情形: $H^2 z = 0$.此时,由于

$$\|Hz\|^2 = (Hz, Hz) = (H^2 z, z) = 0,$$

推出 $Hz = 0$.

现对 m 作归纳.记 $w \equiv H^{m-2} z$,于是从

$$H^2 w = H^2 H^{m-2} z = H^m z = 0,$$

并根据 $m = 2$ 情形的结论,有 $Hw = 0$,即 $H^{m-1} z = 0$.这蕴涵着:若成立

$$H^{m-1} z = 0 \Rightarrow Hz = 0,$$

则成立

$$H^m z = 0 \Rightarrow Hz = 0.$$

这样就完成了归纳法的步骤.

(3) 假定 λ 和 μ 是 H 的特征值,$\lambda \neq \mu$,相应的特征向量分别是 x 和 y:

$$Hx = \lambda x, \quad Hy = \mu y,$$

于是

$$(Hx, y) = \lambda(x, y), \quad (x, Hy) = \mu(x, y)$$

因而根据(2.2)得

$$\lambda(x, y) = \mu(x, y),$$

由此及 $\lambda \neq \mu$ 推出 $(x, y) = 0$,即 x 和 y 正交. □

下一定理是如上定理的推论,它使用 **3.1.3** 中的矩阵形式表达实二次型,说明可以借助实等距变换实现对角化.

3.2.4 定理 设 $H \in \mathbb{R}^{n\times n}$ 是实自伴矩阵(即实对称矩阵),则存在实正交矩阵 M ,即满足 $M^{\mathrm{T}} M = I$,使得

$$M^{\mathrm{T}} HM = D,$$

其中 D 是以 H 的 n 个(按重数计)特征值为对角元素的对角矩阵.

证 设 x 是 H 的特征向量,满足

$$Hx = \lambda x,$$

依 **3.2.3** 中的(1),特征值 λ 是实数.于是,从 H 是实矩阵推出 x 的实部和虚部也是 H 的特征向量.由此可知,在 H 的属于特征值 λ 的所有真特征向量和广义特征向量生成的特征空间 $\mathscr{N}_\lambda$ 中,可以选取由实特征向量组成的正交基.因为依 **3.2.3** 中的(3),属于不同特征值的特征向量是正交的,所以 Euclid 空间 $\mathbb{C}^n$ 存在由实特征向量组成的正交基 $x^{(1)},\cdots,x^{(n)}$,它们分别相应于 H 的全为实数的特征值 $\lambda_1,\cdots,\lambda_n$.$\mathbb{C}^n$ 中每个向量 y 可表示为这些特征向量的线性组合:

$$y = \sum_{i=1}^{n} z_i x^{(i)}, \tag{2.5}$$

当 y 是实向量时,$z=(z_1,\cdots,z_n)^{\mathrm{T}}$ 也是实的.显然

$$\|y\|^2 = \sum_{i=1}^{n} z_i^2 = \|z\|^2, \tag{2.6}$$

即新向量 z 和原向量 y 两者的长度相等;而且

$$Hy = \sum_{i=1}^{n} z_i \lambda_i x^{(i)}. \tag{2.7}$$

因此,形如(1.6)的二次型 q 有表示式

$$q(y) \equiv (y, Hy) = \sum_{i=1}^{n} \lambda_i z_i^2,$$

这表明所引入的新向量 z 使二次型 q 对角化.至此,联合 **3.1.3**,便得所要证明的结论. □

3.2.5 定理 设 $H=[h_{ij}]\in\mathbb{C}^{n\times n}$.下列各条件等价:

(1) H 是自伴矩阵.

(2) $x^* Hx \in \mathbb{R}, \forall x \in \mathbb{C}^n$.

(3) H 是正规矩阵而且所有特征值是实数.

(4) 对于任意的 $S \in \mathbb{C}^{n\times n}$, S^*HS 是自伴矩阵.

证 (1)$\Leftrightarrow$(2): 若 $H^* = H$, 则对于任何 $x \in \mathbb{C}^n$,

$$\overline{\left(x^*Hx\right)} = \left(x^*Hx\right)^* = x^*H^*x = x^*Hx,$$

表明 $x^*Hx \in \mathbb{R}$.

反之,若条件(2)成立,则对于任何 $x, y \in \mathbb{C}^n$,

$$(x+y)^*H(x+y) = \left(x^*Hx + y^*Hy\right) + \left(x^*Hy + y^*Hx\right) \in \mathbb{R}.$$

由此,及依假设知 $x^*Hx + y^*Hy \in \mathbb{R}$,推出

$$x^*Hy + y^*Hx \in \mathbb{R}.$$

特别地,有

$$h_{kj} + h_{jk} = e_k^*He_j + e_j^*He_k \in \mathbb{R}$$

和

$$-ih_{kj} + ih_{jk} = (ie_k)^*He_j + e_j^*H(ie_k) \in \mathbb{R}.$$

于是

$$\operatorname{Im} h_{kj} = -\operatorname{Im} h_{jk}, \quad \operatorname{Re} h_{kj} = \operatorname{Re} h_{jk},$$

从而 $h_{kj} = \overline{h_{jk}}$. 因 j,k 是任意的,故 $A^* = A$.

(1)$\Leftrightarrow$(3): 若 $H^* = H$, 则

$$H^*H = H^2 = HH^*,$$

即 H 是正规矩阵;**3.2.3** 已证明 H 的特征值是实数.

反之,若 H 是正规矩阵,则依 **2.4.4**, H 酉等价于对角矩阵:

$$H = U\Lambda U^*, \quad \Lambda \equiv \operatorname{diag}(\lambda_1, \cdots, \lambda_n),$$

其中 $U \in \mathbb{C}^{n\times n}$ 是酉矩阵, $\lambda_1, \cdots, \lambda_n \in \lambda(H)$. 如果 $\lambda_1, \cdots, \lambda_n \in \mathbb{R}$, 那么

$$H^* = \left(U\Lambda U^*\right)^* = U\Lambda^*U^* = U\Lambda U^* = H.$$

(1)$\Leftrightarrow$(4): 若 $H^* = H$, 则对于任意的 $S \in \mathbb{C}^{n\times n}$,

$$\left(S^*HS\right)^* = S^*H^*S = S^*HS,$$

故 S^*HS 是自伴矩阵.

反之,条件(4)显然蕴涵 H 是自伴矩阵,只须取 $S=I$. □

3.2.6 定理 设 $A\in\mathbb{C}^{n\times n}$,下列各条件等价:

(1) A 相似于一个矩阵 $B\in\mathbb{R}^{n\times n}$.

(2) A 相似于 A^*.

(3) A 通过自伴相似变换相似于 A^*.

(4) $A=HK$,其中 $H,K\in\mathbb{C}^{n\times n}$ 是自伴矩阵而且至少有一个是非奇异的.

(5) $A=HK$,其中 $H,K\in\mathbb{C}^{n\times n}$ 是自伴矩阵.

证 (1)$\Leftrightarrow$(2): 若条件(1)成立,则依 **1.5.17**, 并依 **1.6.20**,存在非奇异矩阵 $S,T\in\mathbb{C}^{n\times n}$,使得

$$S^{-1}AS=B=T^{-1}B^{\mathrm{T}}T=T^{-1}B^*T=T^{-1}S^*A^*\left(S^{-1}\right)^*T,$$

从而

$$A^*=\left(ST^{-1}S^*\right)^{-1}A\left(ST^{-1}S^*\right),$$

即 A 和 A^* 相似.

反之,若条件(2)成立,则 A 和 A^* 有相同的Jordan标准形 J.另一方面,依 **1.6.20**, A 和 A^{T} 是相似的.这意味着 J 相似于 $\overline{J}$.因此,对于 J 中的每个 Jordan 块 $J_k(\lambda)$,在 $\overline{J}$ 中存在相应的同样大小的 Jordan 块 $J_k(\lambda)$.这样,若 λ 为实数,是不成问题的;但若 λ 为复数,则 A 的分别相应于特征值 λ 和 $\overline{\lambda}$ 的 Jordan 块必匹配成对出现,由此推出 J 必相似于一个实矩阵.

(2)$\Rightarrow$(3): 设 $S^{-1}AS=A^*$,则

$$T^{-1}AT=A^*,\quad \forall T=\alpha S,\alpha=re^{i\theta}\in\mathbb{C},r\neq 0.$$

于是

$$AT=TA^*,\quad \text{或等价地 } AT^*=T^*A^*.$$

两式相加,得

$$A\left(T+T^*\right)=\left(T+T^*\right)A^*.$$

从而,如果$T+T^*$是非奇异的, A通过自伴矩阵$T+T^*$相似于A^*.
注意到

$$
\begin{aligned}
&T \text{ 和 } T+T^* \text{ 非奇异} \\
&\quad \Leftrightarrow T^{-1}(T+T^*)=I+T^{-1}T^* \text{ 非奇异} \\
&\quad \Leftrightarrow -1 \notin \lambda(T^{-1}T^*),
\end{aligned}
$$

以及

$$T^{-1}T^* = e^{-2i\theta}S^{-1}S^*,$$

故只须选取$\theta \in [0,2\pi)$,使得

$$-e^{2i\theta} \notin \lambda(S^{-1}S^*).$$

(3)$\Rightarrow$(4): 设

$$R^{-1}AR = A^*,$$

其中$R \in \mathbb{C}^{n\times n}$非奇异而且是自伴的,则

$$A = R(A^*R^{-1}).$$

因为

$$(A^*R^{-1})^* = R^{-1}A = A^*R^{-1}$$

所以A是非奇异自伴矩阵R和自伴矩阵A^*R^{-1}之积.

(4)$\Rightarrow$(2): 设条件(4)成立,且H非奇异,则

$$H^{-1}AH = KH = (HK)^* = A^*,$$

故(2)成立.当K非奇异时,论证是类似的.

(4)$\Rightarrow$(5): 显然.

(5)$\Rightarrow$(1): 设条件(5)成立,如果H和K都是奇异的,考虑

$$U^*AU = (U^*HU)(U^*KU),$$

其中$U \in \mathbb{C}^{n\times n}$是酉矩阵,将$H$对角化成形如

$$U^*HU = \begin{bmatrix} D & 0 \\ 0 & 0 \end{bmatrix} = \tilde{H},$$

$D \in \mathbb{C}^{k\times k}$是非奇异对角矩阵,$k<n$.将$U^*KU$如同$\tilde{H}$进行分块,于是

$$U^*AU=\tilde{H}(U^*KU)=\begin{bmatrix}D & 0\\ 0 & 0\end{bmatrix}\begin{bmatrix}\hat{K} & *\\ * & *\end{bmatrix}=\begin{bmatrix}D\hat{K} & *\\ 0 & 0\end{bmatrix},$$

其中$D\hat{K}\in\mathbb{C}^{k\times k}$是两个自伴矩阵之积,一个是非奇异的,依(4)和(1)等价,$D\hat{K}$相似于一个实矩阵$B\in\mathbb{C}^{k\times k}$.用$J\in\mathbb{C}^{k\times k}$表示$B$的Jordan标准形,从而$A$相似于一个矩阵$C\in\mathbb{C}^{n\times n}$,$C$形如

$$C=\begin{bmatrix}J & *\\ 0 & 0\end{bmatrix},$$

这是上三角矩阵,其特征值是J的特征值,连同$n-k$个0特征值.由此可见,C的Jordan标准形和J两者之间,相伴任何非零特征值的Jordan块的结构必相同.特别,C的与复特征值相伴的Jordan块必匹配共轭成对出现.由此推出C的Jordan标准形相似于一个实矩阵. □

自伴矩阵H是以满足性质$H=H^*$而定义的.扩展如此概念的途径之一是考察满足A相似于A^*的矩阵$A\in\mathbb{C}^{n\times n}$.上述定理按几种方式勾画了此类矩阵.

3.2.7 定理 设$A=[a_{ij}]\in\mathbb{C}^{n\times n}$是严格对角优势的自伴矩阵,而且所有对角元素是正的.则A的所有特征值是正的.

证 从**3.2.3**和**2.2.3**直接推出. □

3.3 正交投影和单位分解

3.3.1 引言 现在进一步考察谱定理**3.2.3**,它可以另用这样一种说法:对于n维Euclid空间X到X的自伴映射H而言,整个空间X可以分解为两两正交的特征空间的直和:

$$X = \mathscr{N}^{(1)} \oplus \cdots \oplus \mathscr{N}^{(m)}, \tag{3.1}$$

其中**特征空间** $\mathscr{N}^{(i)}$ 由 H 具有的实特征值 λ_i 的特征向量全体生成，而且

$$\lambda_i \neq \lambda_j (i \neq j), \quad m \leq n.$$

换言之，X 中每一向量 x 可唯一地分解成

$$x = x^{(1)} + \cdots + x^{(m)}, \quad x^{(i)} \in X^{(i)}, \quad i = 1, \cdots, m. \tag{3.2}$$

由此，有

$$Hx = \lambda_1 x^{(1)} + \cdots + \lambda_m x^{(m)} \tag{3.3}$$

显然，出现在(3.2)中的每个 $x^{(i)}$ 依赖于 x，记其关系为

$$x^{(i)} = P_i(x) \equiv P_i x, \quad i = 1, \cdots, m. \tag{3.4}$$

因为 $\mathscr{N}^{(i)}$ 是 X 的线性子空间，推出 $x^{(i)}$ 线性地依赖于 x，所以 P_i 是线性映射.

从(3.2)，(3.3)和(3.4)，得

$$I = \sum_{i=1}^{m} P_i \tag{3.5}$$

和

$$H = \sum_{i=1}^{m} \lambda_i P_i \tag{3.6}$$

3.3.2 定理 算子 $P_i (i = 1, \cdots, m)$ 有如下性质:

(1) $P_i P_j = 0, \forall i \neq j$.

(2) 等幂性: $P_i^2 = P_i$.

(3) 自伴性: $P_i^* = P_i$.

证 (1)和(2)是 P_i 的定义(3.4)的直接推论.

(3) 对 $x, y \in X$，利用形如(3.2)的分解式，得到

$$\left(P_i x, y\right)=\left(x^{(i)}, y\right)=\left(x^{(i)}, \sum_{j=1}^{m} y^{(j)}\right)=\left(x^{(i)}, y^{(i)}\right).$$

类似地有

$$\left(x, P_i y\right)=\left(x^{(i)}, y^{(i)}\right).$$

因此

$$\left(P_i x, y\right)=\left(x, P_i y\right), \quad \forall x, y \in X,$$

依 **3.2.1**, P_i 是自伴的. □

至此,从 **1.7.16** 和 **1.7.17** 得知,

$$P_i, \quad i=1,\cdots,m$$

是正交投影映射.

形如(3.5)的分解称为对应于自伴映射 H 的**单位分解**. H 的形如(3.6)的分解称为 H 的**谱分解**.

现在,可将 **3.2.3** 中自伴映射的说法改述成以下定理.

3.3.3 定理 设 H 是复 Euclid 空间 X 到 X 的自伴映射，则存在对应于 H 的单位分解和 H 的谱分解. □

谱定理的如此叙述方式对于定义自伴算子的函数非常有用.其最大价值是可以作为代表有限维模式的一种样板.

平方(3.6),并利用 **3.3.2**,得

$$H^2=\sum_{i=1}^{m} \lambda_i^2 P_i.$$

应用归纳法容易证明:

$$H^k=\sum_{i=1}^{m} \lambda_i^k P_i, \quad k=1,2,\cdots \tag{3.7}$$

由此推出,对于任何多项式 p,

$$p(H)=\sum_{i=1}^{n}p(\lambda_i)P_i. \tag{3.8}$$

对于任何函数 f,仿照(3.8)定义 $f(H)$:

$$f(H)\equiv\sum_{i=1}^{m}f(\lambda_i)P_i, \tag{3.9}$$

据此来定义 H 的函数演算.

例如

$$e^{Ht}\equiv\sum_{i=1}^{m}e^{\lambda_i t}P_i.$$

下面的结论是 **3.3.3** 的外延.

3.3.4 定理 设 H 和 K 是一对可交换的自伴矩阵:

$$H^*=H,\quad K^*=K,\quad HK=KH,$$

则它们有一个共同的单位分解

$$I=\sum_{i=1}^{m}P_i,$$

而且成立

$$H=\sum_{i=1}^{m}\lambda_iP_i,\quad K=\sum_{i=1}^{m}\mu_iP_i,$$

其中 $\lambda_1,\cdots,\lambda_m$ 和 $\mu_1,\cdots,\mu_m$ 分别是 H 和 K 的特征值.

证 按照从 **1.6.13** 推出 **3.2.3** 的同样方式,从 **3.3.3** 可以推出本定理.因为 H 和 K 可交换,所以对于 $i=1,\cdots,m$, $H-\lambda_iI$ 和 K 也可交换;由此推出, K 将 $H-\lambda_iI$ 的零空间 $\mathscr{N}_i$ 映射到 $\mathscr{N}_i$.另一方面, K 在 $\mathscr{N}_i$ 上的限制是自伴的.这样, K 可以在 $\mathscr{N}_i$ 上进行谱分解;合并 $i=1,\cdots,m$ 的所有谱分解,便得到 H 和 K 的联合谱分解. □

这一结果可以推广到任何两两可交换的实对称算子的有限集合.

3.4 斜自伴矩阵及其它斜矩阵

3.4.1 定义 $A \in \mathbb{C}^{n\times n}$ 称为**斜自伴矩阵**(skew-selfadjoint matrix)或**斜 Hermite 矩阵**,也称为**反自伴矩阵**(anti-selfadjoint matrix)或**反 Hermite 矩阵**,如果

$$A^* = -A. \tag{4.1}$$

平行地,满足(4.1)的从 Euclid 空间 X 到 X 的线性映射 A 称为**斜(或反)自伴算子**或**斜(或反)Hermite 算子**.

斜自伴矩阵有以下明显的性质(证明留作练习):

(1) 对于任意的 $A \in \mathbb{C}^{n\times n}$, $A - A^*$ 是斜自伴矩阵.

(2) 若 $A, B \in \mathbb{C}^{n\times n}$ 是斜自伴矩阵, $\alpha, \beta \in \mathbb{R}^n$,则 $\alpha A + \beta B$ 是斜自伴矩阵.

(3) 若 $A \in \mathbb{C}^{n\times n}$ 是自伴矩阵,则 $\mathrm{i}A$ 是斜自伴矩阵.

(4) 若 $A \in \mathbb{C}^{n\times n}$ 是斜自伴矩阵,则 $\mathrm{i}A$ 是自伴矩阵.

(5) 若 $A = [a_{ij}] \in \mathbb{C}^{n\times n}$ 是斜自伴矩阵,则 $a_{ii}\left(i = 1, \cdots, n\right)$ 是纯虚数.

(6) 斜自伴矩阵是正规矩阵.

3.4.2 定理 设 $A \in \mathbb{C}^{n\times n}$ 是斜自伴矩阵,则

(1) A 的特征值是纯虚数.

(2) $\mathbb{C}^n$ 中可以选取由 A 的特征向量组成的正交基.

证 对于任何 $M \in \mathbb{C}^{n\times n}$,成立

$$\left(\mathrm{i}M\right)^* = -\mathrm{i}M^*. \tag{4.2}$$

由此,若 A 斜自伴,则 iA 是自伴的.对 iA 适用 **3.2.3**. □

3.4.3 定义　$A\in\mathbb{C}^{n\times n}$ 称为**斜(或反)对称矩阵**,如果

$$A^{\mathrm{T}}=-A. \tag{4.3}$$

斜对称矩阵有以下明显的性质(证明留作练习):

(1) 若 $A=[a_{ij}]\in\mathbb{C}^{n\times n}$ 是斜对称矩阵,则

$$a_{ij}=\begin{cases}0, & i=j,\\ -a_{ji}, & i\neq j,\end{cases}\quad i,j=1,\cdots,n. \tag{4.4}$$

(2) 对于任意的 $A\in\mathbb{C}^{n\times n}$, $A-A^{\mathrm{T}}$ 是斜对称矩阵; A 可以表示成一个对称矩阵 A_1 和一个斜对称矩阵 A_2 之和:

$$A=A_1+A_2;\quad A_1\equiv\frac{A+A^{\mathrm{T}}}{2},\quad A_2\equiv\frac{A-A^{\mathrm{T}}}{2}, \tag{4.5}$$

A_1 称为 A 的**对称部分**, A_2 称为 A 的**斜对称部分**.

3.4.4 定理　设 $A,B\in\mathbb{R}^{n\times n}$, A 是正定矩阵, B 是斜对称矩阵,则 $A+B$ 满足

$$x^{\mathrm{T}}(A+B)x>0,\ \forall x\in\mathbb{R}^n, x\neq 0. \tag{4.6}$$

证　因为 $B^T=-B$,所以对于任何 $x\in\mathbb{R}^n$,有

$$x^{\mathrm{T}}Bx=\left(x^{\mathrm{T}}Bx\right)^{\mathrm{T}}=x^{\mathrm{T}}B^{\mathrm{T}}x=-x^{\mathrm{T}}Bx,$$

推出 $x^TBx=0$.于是,从 A 是正定矩阵,对于 $x\neq 0$,有

$$x^{\mathrm{T}}(A+B)x=x^{\mathrm{T}}Ax+x^{\mathrm{T}}Bx=x^{\mathrm{T}}Ax>0.$$ □

3.4.5 推论　设 $A\in\mathbb{R}^{n\times n}$ 满足

$$x^{\mathrm{T}}Ax>0,\quad \forall x\in\mathbb{R}^n, x\neq 0, \tag{4.7}$$

则 A 的对称部分是正定的. □

注意,$A\in\mathbb{R}^{n\times n}$满足条件(4.7),但因$A$不一定对称,故不一定正定(见正定矩阵的定义 **3.6.1**).

3.4.6 定理 设$S\in\mathbb{R}^{n\times n}$是斜对称矩阵,则$I+S$是非奇异的,而且 **Cayley 变换**

$$T=(I-S)(I+S)^{-1} \tag{4.8}$$

是正交矩阵.

证 从 **3.4.4** 的证明得知,对于任何$y\in\mathbb{R}^n$有$y^{\mathrm{T}}Sy=0$.因此,如果$x\in\mathbb{R}^n$满足

$$(I+S)x=0, \tag{4.9}$$

那么

$$x^{\mathrm{T}}x=x^{\mathrm{T}}x+x^{\mathrm{T}}Sx=x^{\mathrm{T}}(I+S)x=0,$$

推出$x=0$.这就是说,齐次方程组(4.9)只有零解,于是$I+S$是非奇异的.类似地,$I-S$也是非奇异的.

其次,由$S^{\mathrm{T}}=-S$,有

$$T^{\mathrm{T}}=(I-S^{\mathrm{T}})(I+S^{\mathrm{T}})^{-1}=(I+S)(I-S)^{-1},$$

推出$T^{\mathrm{T}}T=I$,从而T是正交矩阵. □

3.4.7 定理 设$A\in\mathbb{R}^{n\times n}$是正交矩阵,而且$I+A$是非奇异的.则$A$可表示成

$$A=(I-S)(I+S)^{-1}, \tag{4.10}$$

其中$S\in\mathbb{R}^{n\times n}$是斜对称矩阵.

证 令

$$S=(I-A)(I+A)^{-1}, \tag{4.11}$$

则

$$\begin{aligned}S^{\mathrm{T}}&=(I-A^{\mathrm{T}})(I+A^{\mathrm{T}})^{-1}=(I-A^{-1})(I+A^{-1})^{-1}\\&=(A-I)A^{-1}A(A+I)^{-1}=-(I-A)(I+A)^{-1}=-S,\end{aligned}$$

故S是斜对称矩阵.

另一方面,容易验证(4.11)等价于(4.10). □

3.4.8 定义 $A\in\mathbb{C}^{n\times n}$ 称为**斜(或反)正交矩阵**,如果

$$A^{-1}=-A^{\mathrm{T}}. \tag{4.12}$$

3.4.9 定理 $A\in\mathbb{C}^{n\times n}$ 为斜正交矩阵的充分必要条件是 $\pm\mathrm{i}A$ 为正交矩阵. 更一般地,对于 $\theta\in\mathbb{R}$, $A^{-1}=\mathrm{e}^{\mathrm{i}\theta}A^{\mathrm{T}}$ 的充分必要条件是 $\pm\mathrm{e}^{\mathrm{i}\theta/2}A$ 为正交矩阵.

证 若 $A^{-1}=\mathrm{e}^{\mathrm{i}\theta}A^{\mathrm{T}}$,则

$$\left(\pm\mathrm{e}^{\mathrm{i}\theta/2}A\right)^{\mathrm{T}}\left(\pm\mathrm{e}^{\mathrm{i}\theta/2}A\right)=\mathrm{e}^{\mathrm{i}\theta}A^{\mathrm{T}}A=\mathrm{e}^{\mathrm{i}\theta}\mathrm{e}^{-\mathrm{i}\theta}A^{-1}A=I.$$

反之,若

$$\left(\pm\mathrm{e}^{\mathrm{i}\theta/2}A\right)^{\mathrm{T}}\left(\pm\mathrm{e}^{\mathrm{i}\theta/2}A\right)=I,$$

则

$$A^{-1}=\left(\pm\mathrm{e}^{\mathrm{i}\theta/2}A\right)^{\mathrm{T}}\left(\pm\mathrm{e}^{\mathrm{i}\theta/2}A\right)A^{-1}=\mathrm{e}^{\mathrm{i}\theta}A^{\mathrm{T}}$$

特别,当 $\theta=\pi$ 时,得

$$A^{-1}=\mathrm{e}^{\mathrm{i}\pi}A^{\mathrm{T}}=-A^{\mathrm{T}}\Leftrightarrow\pm\mathrm{e}^{\mathrm{i}\pi/2}A=\pm\mathrm{i}A \text{ 为正交矩阵},$$

亦即 A 为斜正交矩阵情形的充分必要条件. □

3.5 特征值的变分特性

3.5.1 引言 自伴矩阵的谱分解定理 **3.2.3** 的第一种证明(即指在 **3.2.3** 中给出的证明)基于一般矩阵的谱分解定理,必须应用关于复根存在性的代数学基本定理,然后证明所有根是实的.难免会问:证明自伴矩阵的谱分解能否不借助代数学基本定理呢?回答是肯定的.下面给出的证明在各个方面都优于第一种证明.不仅避免了代数学基本定理,而且在实对称矩阵情形还避免了使用复数.它给出的特征值的变分特性对特征值布局的估计非常有用.最重要的是该证明可

以沿用于无限维空间.

3.5.2 定理 3.2.3 的第二种证明 先作点分析,以便比较自然地引出正式的证明.假定已知$\mathbb{C}^n$有由$H\in\mathbb{C}^{n\times n}$的特征向量组成的正交基,并且$H$的特征值全体是如下从小到大排列的实数:

$$\lambda_1\le\lambda_2\le\cdots\le\lambda_n,$$

使用形如(2.5)和(2.7)的表达式,可以建立关系式

$$\frac{(x,Hx)}{(x,x)}=\frac{\sum_{i=1}^{n}\lambda_i|z_i|^2}{\sum_{i=1}^{n}|z_i|^2},\quad \forall x\in\mathbb{C}^n, \tag{5.1}$$

于是,显然有

$$\lambda_1=\min_{x\neq0}\frac{(x,Hx)}{(x,x)},\quad \lambda_n=\max_{x\neq0}\frac{(x,Hx)}{(x,x)}. \tag{5.2}$$

而且当x分别取H的相应于λ_1和λ_n的特征向量时达到最小值和最大值.

现在给出正式的证明.考虑定义在$\mathbb{C}^n$上的函数$R\equiv R_H$:

$$R(x)\equiv\frac{(x,Hx)}{(x,x)},\quad \forall x\in\mathbb{C}^n,x\neq0. \tag{5.3}$$

R称为H的 **Rayleigh 商**.注意,(5.1)用到要证明的结论,在这里不能利用它.下面直接推证$\min R(x)$问题有解,而且这个解是H的特征向量;在此基础上,用归纳法推出H有正交特征向量完全集(即由H的特征向量组成的$\mathbb{C}^n$的正交基).

因为H自伴,根据(2.4),R是实值函数;而且,R是x的零次齐次函数,即对任一数c,

$$R(cx)=cR(x),$$

因此为了寻求其最大值和最小值,只须限制在单位球面$\|x\|=1$上搜索.单位球面是Euclid空间$\mathbb{C}^n$中的紧集,而且R在它之上是实值连续函数.所以根据分析基本原理,R在单位球面某点——记作

$x^{(1)}$——取最小值.

设 y 是任一向量, t 是实变量,则 $R\left(x^{(1)}+ty\right)$是两个 t 的二次函数之商.利用 H 的自伴性和内积的斜对称性,则

$$R\left(x^{(1)}+ty\right)=\frac{\left(x^{(1)},Hx^{(1)}\right)+2t\operatorname{Re}\left(y,Hx^{(1)}\right)+t^2(y,Hy)}{\left(x^{(1)},x^{(1)}\right)+2t\operatorname{Re}\left(y,x^{(1)}\right)+t^2(y,y)}.$$

因为 $R\left(x^{(1)}+ty\right)$在 $t=0$ 达到最小值,依照微分学,满足

$$\frac{\mathrm{d}}{\mathrm{d}t}R\left(x^{(1)}+ty\right)\bigg|_{t=0}=2\operatorname{Re}\left(y,Hx^{(1)}\right)-\lambda_1\cdot 2\operatorname{Re}\left(y,x^{(1)}\right)=0,$$

其中

$$\lambda_1\equiv\min R(x)=R\left(x^{(1)}\right)=\left(x^{(1)},Hx^{(1)}\right).$$

注意 $\left\|x^{(1)}\right\|=1$,于是,有

$$2\operatorname{Re}\left(y,Hx^{(1)}-\lambda_1x^{(1)}\right)=0.$$

用 $\mathrm{i}y$ 替代 y 又可得

$$2\operatorname{Im}\left(y,Hx^{(1)}-\lambda_1x^{(1)}\right)=0,$$

推出

$$2\left(y,Hx^{(1)}-\lambda_1x^{(1)}\right)=0,\quad\forall y\in\mathbb{C}^n.$$

这就是说,向量 $Hx^{(1)}-\lambda_1x^{(1)}$ 与$\mathbb{C}^n$ 中所有向量(包括它自己)正交;这样的向量只能是零向量,亦即

$$Hx^{(1)}-\lambda_1x^{(1)}=0,$$

因此, λ_1 是 H 的特征值, $x^{(1)}$ 是相应的特征向量.

继而考虑 $x^{(1)}$ 的正交补 X_1,

$$X_1\equiv\left\{x\in\mathbb{C}^n:\left(x,x^{(1)}\right)=0\right\}.$$

显然, $\dim X_1=\dim X-1$.由于

$$\begin{aligned}\left(Hx,x^{(1)}\right)&=\left(x,Hx^{(1)}\right)\\&=\left(x,\lambda_1x^{(1)}\right)=\lambda_1\left(x,x^{(1)}\right)=0,\quad\forall x\in X_1,\end{aligned}$$

表明 H 限制在 X_1 上时是 X_1 到 X_1 的映射,自然仍是自伴的.于是,

如同在整个空间 X 中那样,在 X_1 上也可以提 $\min R(x)$ 问题.这个问题仍然有解,也就是说,存在某点 $x^{(2)} \in X_1$ 使得 R 在 X_1 中取最小值 $\lambda_2 \equiv R(x^{(2)})$;而且,$\lambda_2(\lambda_1 \le \lambda_2)$ 是 H 的特征值,$x^{(2)}$ 是相应的特征向量.至此,已完成按照基础空间的维数作归纳的步骤,因而完成了 H 有正交特征向量完全集的证明. □

上述论证中,逐次计算按递增顺序排列的特征值,归结为一个极小值问题的序列.下一定理给出任一指定特征值的特性,而无须涉及属于排列在其之前的特征值的特征向量.

3.5.3 Courant-Fischer 定理(极小极大原理) 设 $H \in \mathbb{C}^{n\times n}$ 是自伴矩阵,特征值为 $\lambda_1 \le \cdots \le \lambda_n$,则

$$\lambda_i = \min_{\dim S=i}\max_{x\in S,x\ne 0}\frac{(x,Hx)}{(x,x)}, \tag{5.4}$$

其中 S 是$\mathbb{C}^n$ 线性子空间,最小值是对遍取$\mathbb{C}^n$ 的 i 维线性子空间而言的.

证 设 $x^{(1)},\cdots,x^{(n)}$ 是 H 的分别属于特征值 $\lambda_1,\cdots,\lambda_n$ 的正交特征向量.

根据 1.3.4 的(1),线性方程组

$$(x,x^{(j)})=0, \quad j=1,\cdots,i-1, \tag{5.5}$$

在$\mathbb{C}^n$ 的每个 i 维线性子空间中有非零解.因而,对于任一指定的 i 维线性子空间 S,其中必存在向量

$$\tilde{x}=\sum_{j=1}^{n} z_j x^{(j)} \ne 0,$$

满足(5.5),于是有

$$\max_{x\in S,x\ne 0}\frac{(x,Hx)}{(x,x)} \ge \frac{(\tilde{x},H\tilde{x})}{(\tilde{x},\tilde{x})}=\frac{\lambda_i|z_i|^2+\cdots+\lambda_n|z_n|^2}{|z_i|^2+\cdots+|z_n|^2} \ge \lambda_i .$$

由 S 的任意性推出

$$\lambda_i \le \min_{\dim S=i} \max_{x\in S, x\neq 0} \frac{(x, Hx)}{(x, x)}. \tag{5.6}$$

另一方面,若取 $\tilde{S}$ 为 $x^{(1)},\cdots,x^{(i)}$ 生成的空间,则其中每个向量形如

$$x = \sum_{j=1}^{i} z_j x^{(j)},$$

于是有

$$\min_{\dim S=i} \max_{x\in S, x\neq 0} \frac{(x, Hx)}{(x, x)} \le \max_{x\in \tilde{S}, x\neq 0} \frac{(x, Hx)}{(x, x)}$$

$$= \max_{x\in \tilde{S}, x\neq 0} \frac{\lambda_1 |z_1|^2 + \cdots + \lambda_i |z_i|^2}{|z_1|^2 + \cdots + |z_i|^2} \le \lambda_i. \tag{5.7}$$

联合(5.6)和(5.7)即得(5.4). □

3.5.4 Weyl 定理 设 $H, K \in \mathbb{C}^{n\times n}$ 是自伴矩阵,特征值

$$\lambda_i(H), \quad \lambda_i(K), \quad \lambda_i(H+K), \quad i = 1, \cdots, n$$

分别按递增顺序排列.则

$$\lambda_i(H) + \lambda_1(K) \le \lambda_i(H+K) \le \lambda_i(H) + \lambda_n(K),$$
$$i = 1, \cdots, n. \tag{5.8}$$

证 对于任何 $x \in \mathbb{C}^n, x \neq 0$,有

$$\lambda_1(K) \le \frac{x^* K x}{x^* x} \le \lambda_n(K).$$

因此,依 **3.5.3**,对于 $i = 1, \cdots, n$,有

$$\lambda_i(H+K) = \min_{\dim S=i} \max_{x\in X, x\neq 0} \frac{x^*(H+K)x}{x^* x}$$

$$= \min_{\dim S=i} \max_{x\in S, x\neq 0} \left(\frac{x^* H x}{x^* x} + \frac{x^* K x}{x^* x} \right)$$

$$\geq \min_{\dim S=i} \max_{x\in S, x\neq 0} \left(\frac{x^* Hx}{x^* x} + \lambda_1(K) \right)$$

$$= \lambda_i(H) + \lambda_1(K).$$

类似地可证 $\lambda_i(H+K) \leq \lambda_i(H) + \lambda_n(K)$. □

3.5.5 推论 设 $H \in \mathbb{C}^{n\times n}$ 是自伴矩阵, $K \in \mathbb{C}^{n\times n}$ 是半正定矩阵, H 和 $H+K$ 的特征值分别按递增顺序排列,则

$$\lambda_i(H) \leq \lambda_i(H+K), \quad i = 1,\cdots,n. \tag{5.9}$$

证 利用(5.8)中的下界及 $\lambda_1(K) \geq 0$. □

3.6 正自伴映射和正定矩阵

3.6.1 定义 设 X 是复 Euclid 空间, $H \in \mathscr{L}(X,X)$ 是自伴映射,如果满足

$$(x, Hx) > 0, \quad \forall x \in X, x \neq 0, \tag{6.1}$$

则称 H 为**正映射**,记作 $H>0$;如果满足

$$(x, Hx) \geq 0, \quad \forall x \in X, x \neq 0, \tag{6.2}$$

则称 H 为**非负映射**,记作 $H \geq 0$.

平行地,对于矩阵 $H \in \mathbb{C}^{n\times n}$ 和 $X = \mathbb{C}^n$ 来说,如果满足形如(6.1)的关系式,则称 H 为**正定矩阵**(positive definite matrix); 如果满足形如(6.2)的关系式,则称 H 为**半正定矩阵**(positive semi-definite matrix).

当 $-H$ 为正映射、非负映射、正定矩阵、半正定矩阵时,分别称 H 为**负映射**、**非正映射**、**负定矩阵**、**半负定矩阵**.

正定矩阵和负定矩阵统称为**有定矩阵**.半正定矩阵和半负定矩阵统称为**半定矩阵**.所有非有定且非半定矩阵统称为**不定矩阵**.

下面关于矩阵的结论,容易换用映射的说法.

3.6.2 定理 (1) 单位矩阵 $I \in \mathbb{C}^{n\times n}$ 是正定矩阵.

(2) 若 $H, K \in \mathbb{C}^{n\times n}$ 是正定矩阵,则 $H+K$ 和 cH (c 是任意正数)是正定矩阵.

(3) 若 $H \in \mathbb{C}^{n\times n}$ 是正定矩阵,$Q \in \mathbb{C}^{n\times n}$ 是可逆矩阵,则 Q^*HQ 是正定矩阵.

(4) 当 $H \in \mathbb{C}^{n\times n}$ 是自伴矩阵时成立:H 为正定矩阵当且仅当 H 的特征值全是正的.

(5) $\mathbb{C}^{n\times n}$ 中每个正定矩阵是可逆的.

(6) $\mathbb{C}^{n\times n}$ 中每个正定矩阵有唯一的正平方根,而且仍为正定矩阵.

(7) $\mathbb{C}^{n\times n}$ 中所有正定矩阵构成的集合是所有自伴矩阵构成的空间中的一个开子集.

(8) $\mathbb{C}^{n\times n}$ 中所有正定矩阵构成的集合的边界点是半正定矩阵,但不是正定矩阵.

证 (1)是内积具有正定性的推论.(2)是显然的.

(3) 因 Q 可逆,故 $x \in \mathbb{C}^n, x \neq 0$ 时 $y \equiv Qx \neq 0$.于是,从 H 的正定性有

$$(x, Q^*HQx) = (Qx, HQx) = (y, Hy) > 0,$$

$$\forall x \in \mathbb{C}^n, x \neq 0,$$

因此 Q^*HQ 是正定矩阵.

(4) 设 H 为正定矩阵,λ 是 H 的任一特征值,z 是相应的特征向量,$Hz = \lambda z$,则

$$(z, Hz) = \lambda(z, z). \tag{6.3}$$

注意到 $(z, Hz) > 0$ 和 $(z, z) > 0$,从(6.3)推出 $\lambda > 0$.

反过来,设 H 的特征值 $\lambda_1,\cdots,\lambda_n$ 全是正的, $z^{(1)},\cdots,z^{(n)}$ 分别是相应的特征向量并构成$\mathbb{C}^n$ 的正交基,那么任一 $x\in\mathbb{C}^n$ 可以表示成

$$x=\sum_{i=1}^{n}x_i z^{(i)}, \tag{6.4}$$

于是

$$(x,Hx)=\sum_{i,j=1}^{n}\left(x_i z^{(i)},x_j\lambda_j z^{(j)}\right)=\sum_{i=1}^{n}\lambda_i|x_i|^2 .$$

注意到

$$(x,x)=\sum_{i=1}^{n}|x_i|^2=\|x\|^2,$$

即得不等式

$$(x,Hx)\geq\left(\min_{1\leq i\leq n}\lambda_i\right)\|x\|^2 . \tag{6.5}$$

由此推出,当 $x\neq 0$ 时 $(x,Hx)>0$.因此, H 是正定的.

(5) 零是每个不可逆矩阵的特征值;而根据(4),正定矩阵的特征值全是正的,因此正定矩阵是可逆的.

(6) 设 H 是正定矩阵,利用 H 的特征向量构成的$\mathbb{C}^n$ 的正交基,将任一 $x\in\mathbb{C}^n$ 表示成(6.4)的形式.

定义 $\sqrt{H}$ 为

$$\sqrt{H}x\equiv\sum_{i=1}^{n}x_i\sqrt{\lambda_i}z^{(i)}, \tag{6.6}$$

其中 $\lambda_1,\cdots,\lambda_n$ 是 H 的特征值, $z^{(1)},\cdots,z^{(n)}$ 分别是相应的特征向量.显然,

$$\left(\sqrt{H}\right)^2=H,$$

而且 $\sqrt{H}$ 的特征值为 $\sqrt{\lambda_1},\cdots,\sqrt{\lambda_n}$,相应的特征向量仍为 $z^{(1)},\cdots,z^{(n)}$. $\sqrt{H}$ 称为 H 的**正平方根**.

由于 $\sqrt{H}$ 的所有特征值是正的,根据(4),它是正定矩阵.正平方

根的唯一性很容易证明,留作练习.

(7) 设 H 是正定矩阵,$\lambda = \min\limits_{1\le i\le n}\lambda_i$ 是 H 的最小特征值;又设 N 是任一与 H 距离小于 λ 的自伴矩阵,即

$$\|N - H\| < \lambda ,$$

记 $M = N - H$,于是

$$\|Mx\| < \lambda\|x\|.$$

再根据 Schwarz 不等式 **1.7.5**,对于 $x \ne 0$,

$$|(x, Mx)| \le \|x\|\|Mx\| < \lambda\|x\|^2 .$$

由此及(6.5),对于 $x \ne 0$,

$$\begin{aligned}(x, Nx) &= (x,(H + M)x) \\ &= (x, Hx) + (x, Mx) > \lambda\|x\|^2 - \lambda\|x\|^2 = 0,\end{aligned}$$

因此 N 是正定矩阵.这样,证明了 H 以其自己为中心以 λ 为半径的由自伴矩阵组成的开球(邻域),属于$\mathbb{C}^{n\times n}$ 中正定矩阵全体所构成的集合,从而如此集合是所有自伴矩阵构成的空间中的一个开子集.

(8) 设 K 是$\mathbb{C}^{n\times n}$ 中正定矩阵全体所成之集的一个边界点,则存在正定矩阵序列 $\{H_i\}$:

$$\lim_{i\to\infty} H_i = K ,$$

于是由 Schwarz 不等式推出

$$\lim_{i\to\infty}(x, H_i x) = (x, Kx), \quad \forall x \in \mathbb{C}^n .$$

由于每个 H_i 是正定矩阵,正数序列的极限是非负的,因此 K 必是半正定矩阵.另外,K 不可能是正定矩阵,否则从(7)得知它不是边界点. □

3.6.3 定理 (1) 单位矩阵 $I \in \mathbb{C}^{n\times n}$ 是半正定矩阵.

(2) 若 $H, K \in \mathbb{C}^{n\times n}$ 是半正定矩阵,则 $H + K$ 和 cH (c 是任意正数)是半正定矩阵.

(3) 若 $H\in\mathbb{C}^{n\times n}$ 是半正定矩阵, $Q\in\mathbb{C}^{n\times n}$ 是可逆矩阵, 则 Q^*HQ 是半正定矩阵.

(4) 当 $H\in\mathbb{C}^{n\times n}$ 是自伴矩阵时成立: H 为半正定矩阵当且仅当 H 的特征值全是非负的.

(5) $\mathbb{C}^{n\times n}$ 中每个半正定矩阵有唯一的正平方根,而且仍为半正定矩阵.

证 仿照 **3.6.2** 的相应结论的证明,留作练习. □

3.6.4 定义 设 M 和 N 是复 Euclid 空间 X 到 X 的两个自伴映射. 如果

$$N-M>0,$$

即 $N-M$ 是正映射,则称 M **小于** N 或 N **大于** M .记作 $M<N$ 或 $N>M$.如果

$$N-M\geq 0,$$

即 $N-M$ 是非负映射,则称 M **小于等于** N 或 N **大于等于** M .记作 $M\leq N$ 或 $N\geq M$.

换用矩阵语言: $M,N\in\mathbb{C}^{n\times n}$ 是两个自伴矩阵.如果 $N-M$ 是正定矩阵,则称 M **小于** N 或 N **大于** M .如果 $N-M$ 是半正定矩阵,则称 M **小于等于** N 或 N **大于等于** M .

如此定义的自伴映射的偏序具有实数的自然顺序关系的某些性质.如下三条基本性质是 **3.6.4** 的简单推论:

(1) 可加性. 若 $M_1<N_1,M_2<N_2$,则

$$M_1+M_2<N_1+N_2. \tag{6.7}$$

(2) 传递性. 若 $L<M,M<N$ 则 $L<N$.

(3) 乘法性质. 若 $M<N,Q$ 是可逆的,则

$$Q^*MQ<Q^*NQ. \tag{6.8}$$

此外,还成立实数的自然顺序关系的另外某些性质.例如,下面

的倒数性质.

3.6.5 定理 设 M 和 N 是正映射,满足

$$0 < M < N, \tag{6.9}$$

则

$$M^{-1} > N^{-1}. \tag{6.10}$$

换用矩阵语言:设 $M, N \in \mathbb{C}^{n \times n}$ 是两个正定矩阵,而且 $N - M$ 是正定矩阵.则 $M^{-1} - N^{-1}$ 是正定矩阵.

证 首先考虑 $N = I$ 的情形.根据定义,$M < I$ 即指 $I - M$ 是正映射.依 **3.6.2** 的(4),$I - M$ 的特征值是正的,亦即 M 的特征值小于1.又因 M 是正映射,故 M 的特征值在0和1之间.注意到 M^{-1} 的特征值是 M 的特征值的倒数,所以 M^{-1} 的特征值大于1,亦即 $M^{-1} - I$ 的特征值是正的.于是 $M^{-1} - I$ 是正映射,也就是说,$M^{-1} > I$.

现在转向满足(6.9)的任一 N.依 **3.6.2** 的(6),可作因式分解 $N = R^2, R > 0$.再依 **3.6.2** 的(5),知 R 可逆;利用(6.8),取 $Q = R^{-1}$,从(6.9)推出

$$0 < R^{-1}MR^{-1} < R^{-1}NR^{-1} = I.$$

由此,应用关于 $N = I$ 的情形的已证结果,得

$$\left(R^{-1}MR^{-1}\right)^{-1} = RM^{-1}R > I.$$

再次利用(6.8),取 $Q = R^{-1}$,推出

$$M^{-1} > R^{-1}IR^{-1} = R^{-2} = N^{-1}. \qquad \square$$

关于自伴矩阵偏序的进一步讨论见 **10.13**.

3.6.6 定理 若 $A \in \mathbb{C}^{n \times n}$ 是对角元素为正实数的自伴矩阵,而且是严格对角优势或不可约弱严格对角优势的,则 A 是正定矩阵.

证 这是 **2.2.3** 的直接推论. $\square$

3.6.7 定理 设 $A\in\mathbb{C}^{n\times n}$ 是半正定矩阵, $A=0$ 的充分必要条件是 $\mathrm{tr}A=0$.

证 依 **3.2.3**, A 酉等价于对角矩阵,

$$A=U\Lambda U^*,$$

其中 $\Lambda=\mathrm{diag}(\lambda_1,\cdots,\lambda_n)$, $U=\begin{bmatrix}u^{(1)} & \cdots & u^{(n)}\end{bmatrix}$ 是酉矩阵, $u^{(i)}$ 是 A 的属于特征值 λ_i 的特征向量. 依 **3.6.3**, $\lambda_i\ge 0$, $i=1,\cdots n$. 再依 **1.6.6**,

$$\mathrm{tr}A=\lambda_1+\cdots+\lambda_n\ge 0.$$

如果 $A=0$, 显然 $\mathrm{tr}A=0$.

反之, 如果 $\mathrm{tr}A=0$, 则

$$\lambda_1=\cdots=\lambda_n=0,$$

于是 $A=U\Lambda U^*=U0U^*=0$. □

3.6.8 定理 设 $A\in\mathbb{C}^{n\times n}$, 则

$$\|A\|_2=\left(\rho\left(A^*A\right)\right)^{1/2}. \tag{6.11}$$

证 A^*A 是自伴矩阵, 且因

$$x^*\left(A^*A\right)x=\|Ax\|_2^2\ge 0,\quad \forall x\in\mathbb{C}^n,$$

故 A^*A 还是半正定的. 依 **3.2.3** 和 **3.6.3**, 存在由 A^*A 的特征向量组成的 $\mathbb{C}^n$ 的正交基 $x^{(1)},\cdots,x^{(n)}$:

$$A^*Ax^{(i)}=\nu_i x^{(i)},\quad i=1,\cdots,n,$$

其中 $0\le\nu_1\le\cdots\le\nu_n$, ν_i 是对应 $x^{(i)}$ 的特征值, $i=1,\cdots,n$. 由于

$$\left(x^{(i)},x^{(j)}\right)=\begin{cases}1, & i=j,\\ 0, & i\ne j,\end{cases}\quad 1\le i,j\le n,$$

以及对于任意的 $x\in\mathbb{C}^n$, $x\ne 0$, 有表示式

$$x=\sum_{i=1}^{n}c_i x^{(i)},$$

因而

$$\left(\frac{\|Ax\|_2}{\|x\|_2}\right)^2=\frac{x^*A^*Ax}{x^*x}=\frac{\sum_{i=1}^{n}|c_i|^2\nu_i}{\sum_{i=1}^{n}|c_i|^2}.$$

由此推出

$$0\le\nu_1\le\left(\frac{\|Ax\|_2}{\|x\|_2}\right)^2\le\nu_n.$$

注意到当 $x=x^{(n)}$ 时,右边成立等号,最后有

$$\|A\|_2^2=\sup_{x\neq0}\left(\frac{\|Ax\|_2}{\|x\|_2}\right)^2=\nu_n=\rho(A^*A).\qquad\square$$

3.6.9 推论 设 $A\in\mathbb{C}^{n\times n}$ 是自伴矩阵.则

$$\|A\|_2=\rho(A)\tag{6.12}$$

而且,若 $p(t)$ 是实多项式,则

$$\|p(A)\|_2=\rho(p(A)).\tag{6.13}$$

证 因 $A^*=A$,有

$$\|A\|_2^2=\rho(A^*A)=\rho(A^2)=(\rho(A))^2,$$

故(6.12)成立.

当 $p(t)$ 是实多项式时,显然 $p(A)$ 也是自伴矩阵,从而(6.13)成立. $\square$

3.7 自伴矩阵的对称积

3.7.1 定义 设 $A,B\in\mathbb{C}^{n\times n}$. $AB+BA$ 称为 A 和 B 的**对称积**.

两个自伴矩阵 A 和 B 之积(复合) AB 或 BA 一般不是自伴的，但是它们的对称积 $S \equiv AB + BA$ 必是自伴的.此时,与对称积相应的二次型有

$$(x, Sx) = (x, ABx) + (x, BAx) = (Ax, Bx) + (Bx, Ax).$$

进而,在实的情形下,

$$(x, Sx) = 2(Ax, Bx).$$

此式表明,两个正定矩阵 A 和 B 的对称积不一定是正定的;条件

$$(x, Ax) > 0 \quad 和 \quad (x, Bx) > 0$$

分别意味着"向量对" x 与 Ax 的夹角和"向量对" x 与 Bx 的夹角均小于 $\pi/2$,然而这些限制不能避免出现 Ax 与 Bx 的夹角大于 $\pi/2$ —— $(x, Sx) < 0$ ——的情形.

3.7.2 定理 设 $A, B \in \mathbb{C}^{n \times n}$ 是自伴矩阵,并具有如下性质:

(1) A 是正定矩阵,

(2) 对称积 $S \equiv AB + BA$ 是正定矩阵,

则 B 是正定矩阵.

证 定义

$$B(t) \equiv B + tA.$$

考虑 A 和 $B(t)$ 的对称积

$$S(t) \equiv AB(t) + B(t)A = AB + BA + 2tA^2 = S + 2tA^2,$$

显然,当 $t \geq 0$ 时, $S(t)$ 是正定的.因为

$$(x, B(t)x) = (x, Bx) + t(x, Ax),$$

A 是正定的,根据(6.5),

$$(x, Ax) \geq a\|x\|^2,$$

其中 $a > 0$ 是 A 的最小特征值;另一方面,根据 Schwarz 不等式 **1.7.5**,有

$$|(x, Bx)| \leq \|x\|\|Bx\| \leq \|B\|\|x\|^2,$$

所以

$$(x, B(t)x) \geq (ta - \|B\|)\|x\|^2,$$

这表明当 $t > \|B\|/a$ 时 $B(t)$ 是正定的.

由于 $B(t)$ 连续依赖于 t,假若 $B = B(0)$ 是非正定的,则必存在某个 t_0,

$$0 < t_0 \leq \|B\|/a,$$

使得 $B(t_0)$ 处在正定矩阵全体构成的集合的边界上.依 **3.6.2** 的(8)和(4)以及 **3.6.3** 的(4), $B(t_0)$ 是半正定的但不是正定的,其特征值是非负的而且至少有一个为零.因此存在非零向量(对应零特征值的特征向量) y,使得

$$B(t_0)y = 0,$$

从而得出

$$(y, S(t_0)y) = (Ay, B(t_0)y) + (B(t_0)y, Ay) = 0,$$

这与 $S(t_0)$ 的正定性矛盾,所以 B 是正定矩阵. □

下面将给出上述定理的一个推论,它要用到如下有广泛用处的微积分引理.

3.7.3 引理 设 $A(t)$ 是实变量 t 的可微函数,其值是自伴映射;于是其导数 $(\mathrm{d}/\mathrm{d}t)A(t)$ 也是自伴的.假定 $(\mathrm{d}/\mathrm{d}t)A(t)$ 是正的,那么 $A(t)$ 是增函数:

$$A(s) < A(t), \quad \forall s < t. \tag{7.1}$$

当视 $A(t)$ 为矩阵时,引理是说:若 $(\mathrm{d}/\mathrm{d}t)A(t)$ 是正定矩阵,则对于任何 $s < t$, $A(t) - A(s)$ 是正定矩阵.

证 设 x 是与 t 无关的任一非零向量,由于 $(\mathrm{d}/\mathrm{d}t)A(t)$ 是正映射,有

$$\frac{\mathrm{d}}{\mathrm{d}t}(x, A(t)x) = \left(x, \frac{\mathrm{d}}{\mathrm{d}t}A(t)x\right) > 0.$$

因此,根据普通微积分, $(x, A(t)x)$ 是 t 的增函数:

$$(x, A(s)x) < (x, A(t)x), \quad \forall s < t.$$

这蕴涵 $A(t) - A(s) > 0, \forall t > s$,即(7.1)成立. □

3.7.4 定理 设 M 和 N 是正映射,满足

$$0 < M < N, \tag{7.2}$$

则

$$\sqrt{M} < \sqrt{N}. \tag{7.3}$$

换一种说法:若 M 和 N 是正定矩阵,而且 $N - M$ 是正定的,则 $\sqrt{N} - \sqrt{M}$ 是正定的.

证 定义

$$A(t) \equiv M + t(N - M),$$

因为

$$A(0) = M \quad 和 \quad A(1) = N$$

都是正的;又由(7.2),当 $0 < t < 1$ 时,

$$(\mathrm{d}/\mathrm{d}t)A(t) = N - M$$

是正的,而且

$$A(t) = (1 - t)M + tN$$

是两个正映射之和,$A(t)$ 本身也是正的.所以,可以定义

$$R(t) \equiv \sqrt{A(t)}, \quad 0 \le t \le 1. \tag{7.4}$$

不难证明,作为可微正函数 $A(t)$ 的平方根的 $R(t)$ 是可微的.平方(7.4)中的等式得

$$A(t) = [R(t)]^2.$$

再对 t 微分得

$$\frac{\mathrm{d}}{\mathrm{d}t}A(t) = \left[\frac{\mathrm{d}}{\mathrm{d}t}R(t)\right]R(t) + R(t)\left[\frac{\mathrm{d}}{\mathrm{d}t}R(t)\right], \tag{7.5}$$

这样,$R(t)$ 和 $(\mathrm{d}/\mathrm{d}t)R(t)$ 的对称积是 $(\mathrm{d}/\mathrm{d}t)A(t)$.

由于 $(\mathrm{d}/\mathrm{d}t)A(t) = N - M$ 是正的;同样,$R(t)$ 是正的.所以,依 **3.7.2**,$(\mathrm{d}/\mathrm{d}t)R(t)$ 在区间 $[0,1]$ 上是正的.由此并依 **3.7.3**, $R(t)$ 是 t

的增函数;特别,

$$\sqrt{M}=\sqrt{A(0)}=R(0)<R(1)=\sqrt{A(1)}=\sqrt{N}. \qquad \square$$

3.8 Gram 矩阵

3.8.1 定义 设 $f_1,\cdots,f_n$ 是 Euclid 空间中的一组有序向量,矩阵 G 具有元素

$$(G)_{ij}=(f_j,f_i) \quad i,j=1,\cdots,n, \tag{8.1}$$

G 称为向量组 $f_1,\cdots,f_n$ 的 **Gram 矩阵**.

3.8.2 定理 (1) 每个 Gram 矩阵是半正定的.

(2) 线性无关向量组的 Gram 矩阵是正定的.

(3) 每个正定矩阵可以表示成 Gram 矩阵.

证 与元素形如(8.1)的 Gram 矩阵 G 相对应的二次型可以表示成

$$(x,Gx)=\sum_{i,j=1}^{n}x_i\overline{G_{ij}x_j}=\sum_{i,j=1}^{n}(f_i,f_j)x_i\overline{x_j}$$

$$=\left(\sum_{i=1}^{n}x_if_i,\sum_{j=1}^{n}x_jf_j\right)=\left\|\sum_{i=1}^{n}x_if_i\right\|^2, \tag{8.2}$$

由此直接推出结论(1)和(2).

为了证明(3),设矩阵 H 是正定矩阵,在$\mathbb{C}^n$中定义非标准内积$(\bullet,\bullet)_H$如下:

$$(x,y)_H\equiv(x,Hy), \quad \forall x,y\in\mathbb{C}^n, \tag{8.3}$$

其中$(\bullet,\bullet)$是标准内积.单位向量组

$$f_i=e_i, \quad i=1,\cdots,n$$

的 Gram 矩阵的元素是

$$(e_i,e_j)_H=(e_i,He_j)=(H)_{ij},\quad i,j=1,\cdots,n. \qquad \square$$

3.8.3 例 (1) 考虑在区间$[0,1]$上的实值连续函数组成的 Euclid 空间,取内积

$$(f,g)=\int_0^1 f(t)g(t)\mathrm{d}t,$$

向量组

$$f_i(t)=t^{i-1},\quad i=1,\cdots,n$$

的 Gram 矩阵的元素是

$$(G)_{ij}=\frac{1}{i+j-1},\quad i,j=1,\cdots,n. \tag{8.4}$$

(2) 取内积

$$(f,g)=\int_0^{2\pi} f(\theta)\overline{g(\theta)}w(\theta)\mathrm{d}\theta, \tag{8.5}$$

其中 w 是某个给定的正实函数.向量组

$$f_j(\theta)=\mathrm{e}^{\mathrm{i}j\theta},\quad j=-n,\cdots,n$$

的$(2n+1)\times(2n+1)$Gram 矩阵的元素是

$$(G)_{kj}=c_{k-j},\quad k,j=1,\cdots,2n+1,$$

其中

$$c_m=\int_0^{2\pi} w(\theta)\mathrm{e}^{-\mathrm{i}m\theta}d\theta,\quad m=-2n,\cdots,2n.$$

注意,这里$\mathrm{i}=\sqrt{-1}$.

3.9 广义 Rayleigh 商

3.9.1 定义 设$H,M\in\mathbb{C}^{n\times n}$,形如

$$R_{H,M}(x) \equiv \frac{(x,Hx)}{(x,Mx)}, \quad \forall x \in \mathbb{C}^n \tag{9.1}$$

的函数 $R_{H,M}$ 称为矩阵 H 和 M 的**广义 Rayleigh 商**.

取 $M = I$,便回到原先(5.3)定义的 Rayleigh 商 R_H.现在对广义 Rayleigh 商提出前面对原先 Rayleigh 商讨论过的同样问题:极小化 $R_{H,M}$,亦即

求 $\tilde{x} \neq 0, \tilde{x} \in \mathbb{C}^n$,使得

$$R_{H,M}(\tilde{x}) = \min_{x \in C^n, (x,Mx) \neq 0} R_{H,M}(x). \tag{9.2}$$

下面关于广义 Rayleigh 商的若干定理不难应用此前的结果给以证明,可以作为练习.

3.9.2 定理 设 $H, M \in \mathbb{C}^{n \times n}$ 是自伴矩阵,而且 M 是正定的,则

(1) 极小值问题(9.2)有解 $\tilde{x}$.

(2) $\tilde{x}$ 满足方程

$$H\tilde{x} = aM\tilde{x}, \tag{9.3}$$

其中 $a \equiv R_{H,M}(\tilde{x})$.

(3) 带约束极小值问题

$$\min_{(x,M\tilde{x})=0} R_{H,M}(x), \tag{9.4}$$

有解 $\hat{x} \neq 0, \hat{x} \in \mathbb{C}^n$.

(4) $\hat{x}$ 满足方程

$$H\hat{x} = bM\hat{x}, \tag{9.5}$$

其中 $b = R_{H,M}(\hat{x})$. □

3.9.3 定理 设 $H, M \in \mathbb{C}^{n \times n}$ 是自伴矩阵,而且 M 是正定的,则存在 $\mathbb{C}^n$ 的基 $x^{(1)}, \cdots, x^{(n)}$,满足

$$Hx^{(i)} = \lambda_i M x^{(i)}, \quad i = 1, \cdots, n \tag{9.6}$$

和

$$\left(x^{(i)}, Mx^{(j)}\right)=0, \quad \forall i \neq j, \tag{9.7}$$

其中$\lambda_1,\cdots,\lambda_n$是实数. □

对于$\lambda_1,\cdots,\lambda_n$有类似于**3.5.3**的结果.

3.9.4 定理 设$H,M\in\mathbb{C}^{n\times n}$是自伴矩阵,而且$M$是正定的.则$M^{-1}H$的所有特征值是实的.若$H$也是正定的,则$M^{-1}H$的所有特征值是正的. □

这一有用结果是**3.9.3**的直接推论.

3.10 正定矩阵的行列式

3.10.1 定理 每个正定矩阵的行列式是正的.

证 依**1.6.6**,矩阵的行列式等于其所有特征值之积.再依**3.6.2**,正定矩阵的特征值全为正的,故它们之积是正的. □

3.10.2 定理 设$A,B\in\mathbb{R}^{n\times n}$是对称正定的,则

$$\det(tA+(1-t)B)\geq(\det A)^t(\det B)^{1-t},$$

$$\forall t\in[0,1]. \tag{10.1}$$

证 令$C\equiv B^{-1}A$;依**3.9.4**,C有正特征值$\lambda_1,\cdots,\lambda_n$.将不等式(10.1)的左边改写成

$$\det\left[B\left(tB^{-1}A+(1-t)I\right)\right]=\det B\det(tC+(1-t)I).$$

于是,不等式(10.1)等价于

$$\det(tC+(1-t)I)\geq(\det A)^t(\det B)^{-t}=(\det C)^t.$$

将行列式表示成特征值的乘积,进一步得等价形式:

$$\prod_{i=1}^{n}\left(t\lambda_i+1-t\right)\geq\prod_{i=1}^{n}\lambda_i^t .$$

为此,只须证明

$$t\lambda+\left(1-t\right)\geq\lambda^t,\quad \forall t\in[0,1].$$

此不等式是成立的,因为它可以写成

$$tf(1)+(1-t)f(0)\geq f(t),\quad \forall t\in[0,1],$$

而 $f(t)\equiv\lambda^t$ 是凸函数;当 $t=0$ 或 $t=1$ 时取等号. □

3.10.3 Hadamard 不等式 正定矩阵 $A=[a_{ij}]\in\mathbb{C}^{n\times n}$ 的行列式不超过其对角元素之积:

$$\det A\leq\prod_{i=1}^{n}a_{ii}, \tag{10.2}$$

等号成立当且仅当 A 是对角矩阵.

证 A 正定,有

$$a_{ii}=\overline{a_{ii}}=\left(e_i,Ae_i\right)>0,\quad i=1,\cdots,n,$$

即 A 的对角元素全为正的.定义对角矩阵 $D=\operatorname{diag}\left(d_1,\cdots,d_n\right)$,

$$d_i=\frac{1}{\sqrt{a_{ii}}},\quad i=1,\cdots,n.$$

定义矩阵 B 为

$$B\equiv DAD.$$

显然,B 是正定的,且其对角元素全为1.根据行列式乘法性质,

$$\det B=\det A\det D^2=\frac{\det A}{\prod_{i=1}^{n}a_{ii}}. \tag{10.3}$$

所以(10.2)等价于

$$\det B\leq 1.$$

为了证明这一点,用 $\mu_1,\cdots,\mu_n$ 表示 B 的特征值,因 B 正定,它们全是正的.利用算术-几何平均不等式

$$\prod_{i=1}^{n}\mu_i \le \left(\sum_{i=1}^{n}\mu_i \Big/ n\right)^n ,$$

此不等式可改写成

$$\det B \le \left(\frac{\mathrm{tr}B}{n}\right)^n .$$

注意到 B 的对角元素全为 1, $\mathrm{tr}B = n$, 因此 $\det B \le 1$.

算术-几何平均不等式成立等号当且仅当 $\mu_1 = \cdots = \mu_n = 1$;因依 **3.2.3**, B 是酉可对角化的,故其特征值 $\mu_1 = \cdots = \mu_n = 1$ 等价于存在酉矩阵 U 使得 $U^*BU = I$; 然而 $U^*BU = I$ 又等价于 $B = UIU^* = I$. 所以不等式(10.2)成立等号当且仅当

$$A = D^{-1}BD^{-1} = D^{-1}ID^{-1} = D^{-2} = \mathrm{diag}(a_{11}, \cdots, a_{nn}),$$

这就是说, A 是对角矩阵. □

3.10.4 Hadamard 不等式 设 $A = [a_{ij}] \in \mathbb{C}^{n\times n}$, 则 A 的行列式按模不超过其各行 $a_r^{(1)}, \cdots, a_r^{(n)}$ 的长度之积:

$$|\det A| \le \prod_{i=1}^{n}\|a_r^{(i)}\| = \prod_{i=1}^{n}\left(\sum_{j=1}^{n}|a_{ij}|^2\right)^{1/2} , \qquad (10.4)$$

而且按模不超过其各列 $a_c^{(1)}, \cdots, a_c^{(n)}$ 的长度之积:

$$|\det A| \le \prod_{j=1}^{n}\|a_c^{(j)}\| = \prod_{j=1}^{n}\left(\sum_{i=1}^{n}|a_{ij}|^2\right)^{1/2} , \qquad (10.5)$$

(10.4)的等号成立当且仅当 A 的各行正交;(10.5)的等号成立当且仅当 A 的各行正交.

证 $\det A = 0$ 时,(10.4)显然成立.

现设 $\det A \ne 0$, 令

$$H \equiv [h_{ij}] = AA^* ,$$

则 H 正定,而且

$$h_{ii}=\sum_{j=1}^{n}a_{ij}a_{ji}^{*}=\sum_{j=1}^{n}a_{ij}\overline{a_{ij}}=\sum_{j=1}^{n}\left|a_{ij}\right|^{2}=\left\|a_{r}^{(i)}\right\|^{2},$$

应用 **3.10.3**,

$$\det H\leq\prod_{i=1}^{n}\left\|a_{r}^{(i)}\right\|^{2}.$$

另一方面,

$$\det H=\det A\det A^{*}=\det A\overline{\det A}=\left|\det A\right|^{2}.$$

综合以上两个关系式即得(10.4).注意,H 为对角矩阵当且仅当 A 的各行正交;此时(10.4)的等号成立.

类似地,将 **3.10.3** 应用于 $H^{*}=A^{*}A$ 可推证(10.5). □

此定理当 A 是实矩阵时有明显的几何意义:在边长为 A 的各行或各列的长度 $\left\|a^{(1)}\right\|,\cdots,\left\|a^{(n)}\right\|$ 的所有平行体中,体积最大的是矩形体.

3.10.5 引理 设 $A\in\mathbb{C}^{n\times n}$ 是自伴矩阵,用 $i_{+}(A)$ 表示 A 的正特征值的个数,$\tilde{A}$ 表示 A 的比 A 小一阶的主子矩阵,则

$$i_{+}(A)-1\leq i_{+}(\tilde{A})\leq i_{+}(A).$$

证 从 **3.1.2** 的证明中得知 $i_{+}(A)$ 是 $\mathbb{C}^{n}$ 中满足如下条件的最大维数子空间 S 的维数 $\dim S$:

$$(u,Au)>0,\quad\forall u\in S,u\neq 0.\tag{10.6}$$

$i_{+}(\tilde{A})$ 可类似地加以刻画.考虑 $\mathbb{C}^{n-1}$ 中维数为 $i_{+}(\tilde{A})$ 且满足如下条件的子空间 $\hat{S}$:

$$(\hat{u},A\hat{u})>0,\quad\forall\hat{u}\in\hat{S},\hat{u}\neq 0.\tag{10.7}$$

注意,作用于主子式 $\tilde{A}$ 的向量 $\hat{u}$ 比作用于 A 的向量 u 少一个分量.定义 $\mathbb{C}^{n}$ 的子空间 $\widehat{S}$,其向量 u 的第一个分量皆为零,余下的 $n-1$ 个分量组成的向量 $\hat{u}$ 属于 $\hat{S}$:

$$\widetilde{S} \equiv \left\{u = (0, u_2, \cdots, u_n)^{\mathrm{T}} \in \mathbb{C}^n : \hat{u} = (u_2, \cdots, u_n)^{\mathrm{T}} \in \hat{S}\right\}.$$

显然, $\forall u \in \widetilde{S}$,

$$(u, Au) = (\hat{u}, \tilde{A}\hat{u}). \tag{10.8}$$

于是,根据 i_+ 的特性有

$$i_+(A) \geq \dim \widetilde{S} = \dim \hat{S} = i_+(\tilde{A}).$$

另一方面,反过来从满足(10.6)的最大维数子空间 S 着手,考虑 S 中第一个分量为零的那些向量组成的子空间 $\breve{S}$.显然,

$$\dim \breve{S} \geq \dim S - 1.$$

现在用 $\hat{S}$ 表示 $\mathbb{C}^{n-1}$ 的子空间,它是由 $\breve{S}$ 中的 u 删除第一个分量而得到的向量 $\hat{u}$ 组成的.因为删除的分量是零,(10.8)对所有 $u \in \breve{S}$ 成立,所以从(10.6)推出(10.7),并因此

$$i_+(\tilde{A}) \geq \dim \hat{S} = \dim \breve{S} \geq \dim S - 1 = i_+(A) - 1. \qquad \square$$

3.10.6 定理 设 $A \in \mathbb{C}^{n\times n}$ 是自伴矩阵. A 正定的充分必要条件是

$$\det A^{(i)} > 0, \quad i = 0, 1, \cdots, n-1,$$

其中 $A^{(i)} \equiv A(\{i+1, \cdots, n\})$,即 A 删除前 i 行和列而得到的(后)主子矩阵.

等价地,应用中常把充分必要条件中的 $A^{(i)}$ 改换为前主子矩阵,即 $A^{(i)} \equiv A(\{1, \cdots, i\})$.

证 正定矩阵的所有主子矩阵必是正定的,所以依 **3.10.1**,它们的行列式全是正的.必要性得证.

为了证明充分性,在 **3.10.5** 中用 $-A$ 替换 A,得

$$i_-(A) - 1 \leq i_-(\tilde{A}) \leq i_-(A),$$

其中 $i_-(A)$ 表示 A 的负特征值的个数.因为 $A^{(i+1)}$ 是 $A^{(i)}$ 的比 $A^{(i)}$ 小一阶的主子矩阵,在上述不等式中置 $A = A^{(i)}$,得知 $A^{(i+1)}$ 的负特征值不会多于 $A^{(i)}$,并且最多少一个.

$A^{(i)}$ 的行列式是其特征值之积;因此,如果成立

$$\det A^{(i)} > 0, \quad i = 0,1,\cdots,n-1,$$

那么所有 $A^{(i)}$ 有偶数个负特征值.因为 $A^{(i)}$ 与 $A^{(i+1)}$ 的负特征值的个数最多差一个,推出 $A^{(i)}$ 与 $A^{(i+1)}$ 的负特征值的个数一样多.

注意到 $A^{(n-1)}$ 是 1×1 主子矩阵,没有负特征值.于是,基于上述理由,

$$A^{(n-2)}, A^{(n-3)}, \cdots, A^{(0)} = A,$$

都没有负特征值. □

3.10.7 定理 设 $A \in \mathbb{C}^{n\times n}$ 是自伴矩阵, $\tilde{A}$ 是 A 的 $n-1$ 阶主子矩阵. 则 $\tilde{A}$ 的特征值**隔离** A 的特征值:

$$\lambda_1 \le \mu_1 \le \lambda_2 \le \cdots \le \mu_{n-1} \le \lambda_n, \tag{10.9}$$

其中

$$\lambda_1 \le \cdots \le \lambda_n$$

是 A 的特征值,

$$\mu_1 \le \cdots \le \mu_{n-1}$$

是 $\tilde{A}$ 的特征值.

证 应用 **3.10.5** 于矩阵 $A - cI$, c 是实数,推出超过 c 的 μ_i 的个数不多于超过 c 的 λ_i 的个数,而且至多少一个.这蕴涵了(10.9). □

为了好用,有时把这个定理叙述为:

设 $A \in \mathbb{C}^{n\times n}$ 是自伴矩阵, $\hat{A} \in \mathbb{C}^{(n+1)\times(n+1)}$ 是自伴矩阵而且是 A 的**加边矩阵**:

$$\hat{A} \equiv \begin{bmatrix} A & y \\ y^* & a \end{bmatrix}, \quad a \in \mathbb{R}, \ y \in \mathbb{C}^n,$$

则 A 的特征值隔离 $\hat{A}$ 的特征值.

3.10.8 定理 设 $A \in \mathbb{R}^{n\times n}$ 是对称正定矩阵,则

$$\frac{\pi^{n/2}}{\sqrt{\det A}} = \int_{\mathbf{R}^n} e^{-(x, Ax)} dx \tag{10.10}$$

证 不等式(6.5)保证了(10.10)右边的积分收敛:

$$\int_{\mathbf{R}^n} e^{-(x, Ax)} dx \le \int_{\mathbf{R}^n} e^{-\lambda\|x\|^2} dx < \infty,$$

其中 λ 是 A 的最小特征值.为了计算此积分,应用**3.2.4**,引进新坐标

$$x = My,$$

M 是正交矩阵,其选取使得对角化二次型:

$$(x, Ax) = (My, AMy) = (y, M^* AMy) = \sum_{i=1}^{n} \lambda_i y_i^2, \tag{10.11}$$

这里 $\lambda_1, \cdots, \lambda_n$ 是 A 的特征值.因矩阵 M 是等距的,故它既保体积,又有 $|\det M| = 1$.

现在,将(10.11)代入(10.10)右端的积分,新变量使被积函数成为单变量函数的乘积,因此 n 元积分可以写成 n 个一元积分的乘积:

$$\int_{\mathbf{R}^n} e^{-(x, Ax)} dx = \int_{\mathbf{R}^n} e^{-\sum_{i=1}^{n} \lambda_i y_i^2} dy = \prod_{i=1}^{n} \int_{\mathbf{R}^n} e^{-\lambda_i y_i^2} dy_i. \tag{10.12}$$

进而,对 $i = 1, \cdots, n$,作变量替换

$$z_i = \sqrt{\lambda_i} y_i,$$

得

$$\int_{\mathbf{R}^n} e^{-\lambda_i y_i^2} dy_i = \int_{\mathbf{R}^n} e^{-z_i^2} \frac{dz_i}{\sqrt{\lambda_i}} = \frac{\sqrt{\pi}}{\sqrt{\lambda_i}},$$

于是

$$\int_{\mathbf{R}^n} e^{-(x, Ax)} dx = \prod_{i=1}^{n} \frac{\sqrt{\pi}}{\sqrt{\lambda_i}} = \frac{\pi^{n/2}}{\sqrt{\prod_{i=1}^{n} \lambda_i}}.$$

注意到$\prod_{i=1}^{n}\lambda_i=\det A$,即得(10.10). □

应用此定理可以给出**3.10.2**的另一种证明和**3.10.3**的另一种证明.

3.11 关于自伴矩阵特征值的几个不等式

3.11.1 定理 设$A,B\in\mathbb{R}^{n\times n}$是自伴矩阵(也就是说,$A$和$B$是实对称矩阵);而且,$B-A$是正定的,亦即满足

$$(x,Ax)<(x,Bx),\ \forall x\in\mathbb{R}^n,\ x\neq 0. \tag{11.1}$$

设按递增的顺序排列,A的特征值为

$$\lambda_1\leq\cdots\leq\lambda_n,$$

B的特征值为

$$\mu_1\leq\cdots\leq\mu_n,$$

则

$$\lambda_i<\mu_i,\quad i=1,\cdots,n. \tag{11.2}$$

证 应用极小极大原理**3.5.3**,对任一确定的i,有

$$\lambda_i=\min_{\dim S=i}\max_{x\in S,x\neq 0}\frac{(x,Ax)}{(x,x)},\quad \mu_i=\min_{\dim S=i}\max_{x\in S,x\neq 0}\frac{(x,Bx)}{(x,x)}, \tag{11.3}$$

用$\tilde{S}$表示达到极小极大值μ_i的$\dim\tilde{S}=i$的子空间;用$\tilde{x}$表示在$\tilde{S}$中使$(x,Ax)/(x,x)$达到最大值的向量,并且取$\tilde{x}$是规范化的,$\|\tilde{x}\|=1$,即有

$$(\tilde{x},A\tilde{x})=\frac{(\tilde{x},A\tilde{x})}{(\tilde{x},\tilde{x})}=\max_{x\in\tilde{S},x\neq 0}\frac{(x,Ax)}{(x,x)}.$$

那么,根据(11.3),

$$\lambda_i \leq \left(\tilde{x}, A\tilde{x}\right),$$

而且

$$\left(\tilde{x}, B\tilde{x}\right) \leq \mu_i,$$

于是,由(11.1)即得 $\lambda_i < \mu_i$. □

这一定理及下面结论在数学物理中非常有用.

3.11.2 定理 设 $A, B \in \mathbb{R}^{n\times n}$ 是自伴矩阵(也就是说,A 和 B 是实对称矩阵),而且 A 的特征值为 $\lambda_1 \leq \cdots \leq \lambda_n$,$B$ 的特征值为 $\mu_1 \leq \cdots \leq \mu_n$,则

$$\left|\lambda_i - \mu_i\right| \leq \|A - B\|, \quad i = 1, \cdots, n. \tag{11.4}$$

证 记

$$d \equiv \|A - B\|.$$

容易看出

$$\left(x, \left(B - dI\right)x\right) \leq \left(x, Ax\right) \leq \left(x, \left(B + dI\right)x\right), \quad \forall x \in \mathbb{R}^n.$$

由此,并从 **3.11.1** 推出(11.4). □

3.11.3 定理(Wielandt & Hoffman) 设 $A, B \in \mathbb{R}^{n\times n}$ 是自伴矩阵,而且 A 的特征值为 $\lambda_1 \leq \cdots \leq \lambda_n$,$B$ 的特征值为 $\mu_1 \leq \cdots \leq \mu_n$.则

$$\sum_{i=1}^{n}\left(\lambda_i - \mu_i\right)^2 \leq \|A - B\|_F^2, \tag{11.5}$$

这里 $\|\bullet\|_F$ 是矩阵的 Frobenius 范数(见 **1.8.29**).

证 Frobenius 范数可以表示成

$$\|C\|_F^2 = \mathrm{tr}C^*C, \quad \forall C \in \mathbb{C}^{n\times n}. \tag{11.6}$$

当 C 自伴时,

$$\|C\|_F^2 = \mathrm{tr}C^2.$$

于是不等式(11.5)可以改写成

$$\sum_{i=1}^{n}(\lambda_i-\mu_i)^2\le \mathrm{tr}(A-B)^2 .$$

将其两边展开,并利用迹的线性和可交换性,

$$\sum_{i=1}^{n}\lambda_i^2-2\sum_{i=1}^{n}\lambda_i\mu_i+\sum_{i=1}^{n}\mu_i^2\le \mathrm{tr}A^2-2\mathrm{tr}(AB)+\mathrm{tr}B^2 . \tag{11.7}$$

由于矩阵的迹是其特征值之和,而且 A^2 的特征值是 $\lambda_1^2,\cdots,\lambda_n^2$,有

$$\sum_{i=1}^{n}\lambda_i^2=\mathrm{tr}A^2,\quad \sum_{i=1}^{n}\mu_i^2=\mathrm{tr}B^2 .$$

因此不等式(11.7)缩简为

$$\sum_{i=1}^{n}\lambda_i\mu_i\ge \mathrm{tr}(AB), \tag{11.8}$$

这样,问题归结为证明不等式(11.8).

固定 B.考虑特征值为 $\lambda_1,\cdots,\lambda_n$ 的所有自伴矩阵,它们构成整个自伴矩阵空间中的紧支集;在这个紧支集中寻求矩阵 $A_{\max}$ 使(11.8)的右边达最大值.依微分学,$A_{\max}$ 有如下性质:若 $A(t)$ 是值为对称矩阵具有特征值 $\lambda_1,\cdots,\lambda_n$ 的可微函数,且 $A(0)=A_{\max}$,则

$$\frac{\mathrm{d}}{\mathrm{d}t}\mathrm{tr}\big(A(t)B\big)\Big|_{t=0}=0 . \tag{11.9}$$

设 C 是任一斜自伴矩阵,依 **9.5.12** 的(5),对于任意实值 t,e^{Ct} 是酉矩阵.现定义

$$A(t)\equiv \mathrm{e}^{Ct}A_{\max}\mathrm{e}^{-Ct},$$

显然,$A(t)$ 是自伴的,且与 $A_{\max}$ 有相同的特征值.依 **9.5.12** 的(4),

$$\frac{\mathrm{d}}{\mathrm{d}t}\mathrm{e}^{Ct}=C\mathrm{e}^{Ct}=\mathrm{e}^{Ct}C .$$

于是

$$\frac{\mathrm{d}}{\mathrm{d}t}A(t)=\mathrm{e}^{Ct}\left(CA_{\max}-A_{\max}C\right)\mathrm{e}^{-Ct}.$$

将此式代入(11.9),

$$\left.\frac{\mathrm{d}}{\mathrm{d}t}\operatorname{tr}\left(A(t)B\right)\right|_{t=0}=\operatorname{tr}\left[\left(\frac{\mathrm{d}}{\mathrm{d}t}A(t)\right)B\right]_{t=0}$$

$$=\operatorname{tr}\left(CA_{\max}B-A_{\max}CB\right)=0,$$

利用迹的可交换性,上式可改写成

$$\operatorname{tr}\left[C\left(A_{\max}B-BA_{\max}\right)\right]=0. \tag{11.10}$$

从 $A_{\max}$ 和 B 是自伴的,推出 $A_{\max}B-BA_{\max}$ 是斜自伴的,故可选取

$$C=A_{\max}B-BA_{\max},$$

将其代入(11.10)得

$$\operatorname{tr}C^2=0.$$

另一方面,根据(11.6),对于斜自伴的 C,

$$\operatorname{tr}C^2=-\sum_{i,j=1}^{n}\left|c_{ij}\right|^2,$$

因此 $C=0$,从而 $A_{\max}$ 和 B 可交换;这样,$A_{\max}$ 和 B 可同时对角化,对角元素分别是按某种顺序排列的 λ_i 和 μ_i.在这种情形下,

$$\operatorname{tr}\left(A_{\max}B\right)=\sum_{i=1}^{n}\lambda_{p_i}\mu_i, \tag{11.11}$$

其中 $p_1,\cdots,p_n$ 是 $1,\cdots,n$ 的某个置换.

不难证明,(11.11)的右端当 λ_i 和 μ_i 同样以递增顺序排列时达最大.至此,证明了不等式(11.8)对 $A_{\max}$ 成立,因此它对所有以 $\lambda_1,\cdots,\lambda_n$ 为特征值的矩阵 A 成立. □

3.11.4 定理 用 $e_{\min}(C)$ 和 $e_{\max}(C)$ 分别表示自伴矩阵 $C\in\mathbb{C}^{n\times n}$ 的最小特征值和最大特征值.

那么,$e_{\min}$ 是凹函数,也就是说,对任何一对自伴矩阵

$A,B\in\mathbb{C}^{n\times n}$,成立

$$e_{\min}\left(tA+(1-t)B\right)\ge te_{\min}(A)+(1-t)e_{\min}(B),\quad 0<t<1,\tag{11.12}$$

$e_{\max}$ 是凸函数,也就是说,对任何一对自伴矩阵 $A,B\in\mathbb{C}^{n\times n}$,成立

$$e_{\max}\left(tA+(1-t)B\right)\le te_{\max}(A)+(1-t)e_{\max}(B),\quad 0<t<1.\tag{11.13}$$

证 (5.2)表明,

$$e_{\min}(C)=\min_{\|x\|=1}(x,Cx).$$

取

$$C=tA+(1-t)B,$$

令 $\tilde{x}$ 是单位向量,$\|\tilde{x}\|=1$,它使 (x,Cx) 达极小,则

$$\begin{aligned}e_{\min}\left(tA+(1-t)B\right)&=t(\tilde{x},A\tilde{x})+(1-t)(\tilde{x},B\tilde{x})\\&\ge t\min_{\|x\|=1}(x,Ax)+(1-t)\min_{\|x\|=1}(x,Bx)\\&=te_{\min}(A)+(1-t)e_{\min}(B),\end{aligned}$$

(11.12)得证.

现在利用

$$-e_{\max}(C)=e_{\min}(-C),$$

以及(11.12),即可推出(11.13). □

3.12 任意矩阵的表示法

3.12.1 定义 每个矩阵 $C\in\mathbb{C}^{n\times n}$ 可以唯一地分解成一个自伴矩阵与一个斜自伴矩阵之和:

$$C=A+B,\tag{12.1}$$

其中

$$A^* = A, \quad B^* = -B. \tag{12.2}$$

显然,若(12.1)和(12.2)成立,则

$$C^* = A^* + B^* = A - B,$$

因此

$$A = \frac{C + C^*}{2}, \quad B = \frac{C - C^*}{2},$$

A 称为 C 的**自伴部分**, B 称为 C 的**斜自伴部分**.

3.12.2 定理 设 $C \in \mathbb{C}^{n \times n}$ 的自伴部分是正定矩阵,则 C 的特征值有正实部.

证 利用内积的斜对称性和伴随的定义,并根据假设 $C + C^*$ 是正定的,

$$2\operatorname{Re}(Cx, x) = (Cx, x) + \overline{(Cx, x)} = (Cx, x) + (x, Cx)$$

$$= (Cx, x) + (C^* x, x) = ((C + C^*)x, x) \quad \forall x \in \mathbb{C}^n$$

即 (Cx, x) 有正实部.

现在设 x 是 C 的特征向量,而且 $\|x\| = 1$, λ 是相应的特征值,即有 $Cx = \lambda x$,则

$$\lambda = (Cx, x).$$

因此 λ 有正实部. □

3.12.3 定理 设 $C \in \mathbb{C}^{n \times n}$, C 的自伴部分 $C + C^*$ 为正定矩阵的充分必要条件是

$$\left\| (I - cC)(I + \bar{c}C)^{-1} \right\| < 1, \tag{12.3}$$

其中 $c \in \mathbb{C}, \operatorname{Re} c > 0$.

证 C 的自伴部分是正定矩阵时,从 **3.12.2** 推出 $I + \bar{c}C$ 是可逆的.因此,无论讨论充分性还是讨论必要性,矩阵

$$W \equiv (I - cC)(I + \bar{c}C)^{-1} \tag{12.4}$$

都是有意义的.条件(12.3)可以简写成 $\|W\| < 1$.

对于任意的向量$x \in \mathbb{C}^n$,令

$$y \equiv \left(I + \bar{c}C\right)^{-1} x,$$

从(12.4)有

$$\left(I - cC\right)y = Wx,$$

又从y的定义

$$\left(I + \bar{c}C\right)y = x.$$

由于条件$\|W\| < 1$蕴涵

$$\|Wx\|^2 < \|x\|^2, \ \forall x \in \mathbb{C}^n, x \neq 0.$$

借助y可将其表示为

$$\|y - cCy\|^2 < \|y + \bar{c}Cy\|^2, \quad \forall y \in \mathbb{C}^n, y \neq 0.$$

将不等号两边展开,

$$\begin{aligned} &\|y\|^2 + |c|^2\|Cy\|^2 - c(Cy, y) - \bar{c}(y, Cy) \\ &\qquad < \|y\|^2 + |c|^2\|Cy\|^2 + \bar{c}(Cy, y) + c(y, Cy), \\ &\qquad\qquad\qquad \forall y \in \mathbb{C}^n, y \neq 0. \end{aligned}$$

整理后得条件$\|W\| < 1$等价于

$$2(\operatorname{Re} c)((C + C^*)y, y) > 0, \forall y \in \mathbb{C}^n, y \neq 0. \tag{12.5}$$

由于假设$\operatorname{Re} c > 0$,因此条件(12.5)成立——$\|W\| < 1$即(12.3)成立——当且仅当$C + C^*$为正定矩阵. □

换用线性映射说法,复 Euclid 空间X到X的每个线性映射可以唯一地分解成一个自伴映射与一个斜自伴映射之和.

矩阵和线性映射的如此表示法类似于复数写成实部与虚部之和,范数相当于绝对值.

作为例子,设c是具有正实部的复数,则

$$z \mapsto \frac{1 - cz}{1 + \bar{c}z} \equiv w.$$

将复平面的右半平面 $\mathrm{Re}\, z > 0$ 映射成单位圆盘

$$|w| < 1.$$

3.12.3 就是与此类似的结果.

3.13 自伴矩阵多重特征值分析

3.13.1 引言 一般矩阵单重特征值依赖矩阵的基本结论在 **9.5** 中讨论.然而对于多重特征值,问题就要复杂得多.

如果限于考察自伴矩阵,因没有广义特征向量,可以避免广义特征向量增添的困难.但是,即使如此,对于自伴矩阵情形,仍须附加假设才能作出特征值连续依赖矩阵的结果.

看一个简单的 2×2 自伴矩阵值函数的例子.

设

$$A(t) = \begin{bmatrix} b(t) & c(t) \\ c(t) & d(t) \end{bmatrix},$$

并且

$$b(0) = d(0) = 1, \quad c(0) = 0,$$

由于 $A(0) = I$,它以1为重特征值.

而 $A(t)$ 的特征值为

$$\lambda(t) = \frac{b(t) + d(t) \pm \sqrt{(b(t) - d(t))^2 + 4(c(t))^2}}{2}.$$

设特征向量为

$$x(t) = (x_1(t), x_2(t))^{\mathrm{T}}$$

则满足

$$b(t)x_1(t) + c(t)x_2(t) = \lambda(t)x_1(t).$$

因此,可得

$$\frac{x_2(t)}{x_1(t)}=\frac{\lambda(t)-b(t)}{c(t)}=\frac{k(t)+\sqrt{(k(t))^2+4}}{2},$$

其中

$$k(t)=(d(t)-b(t))/c(t).$$

现在,选取

$$k(t)=\sin\frac{1}{t},\quad c(t)=\mathrm{e}^{-\frac{1}{|t|}},$$

且令

$$b(t)\equiv 1,\ d(t)=1+c(t)k(t).$$

显然,$A(t)$的每个元素是无穷可微函数,然而当$t\to 0$时$x_2(t)/x_1(t)$是不连续的.

下面分析一下为了特征向量连续变化需要附加些什么条件.

设$A(t)=[a_{ij}(t)]$是自伴矩阵,其每个元素a_{ij}是实变量t的可微函数.假定在$t=0$,$A(0)$有$k>1$重特征值λ_0.因没有广义特征向量,故$A(0)$属于λ_0的特征向量生成k维空间,记作$\mathscr{N}$.

现在考虑$A(t)$的特征值$\lambda(t)$和相应的特征向量$x(t)$.假定$x(t)$(的每个分量)可微地依赖于t,则可对

$$A(t)x(t)=\lambda(t)x(t)$$

求导,且在$t=0$成立

$$\dot{A}(0)x(0)+A(0)\dot{x}(0)=\dot{\lambda}(0)x(0)+\lambda(0)\dot{x}(0),\tag{13.1}$$

这里用点表示导数,

$$\dot{A}(t)\equiv[\dot{a}_{ij}(t)]=\left[\frac{\mathrm{d}}{\mathrm{d}t}a_{ij}(t)\right],$$

记号$\dot{x}(t)$的含义相仿.用P表示映射至$\mathscr{N}$上的正交投影,依(3.5)和(3.6)及**3.3.2**,推出

$$PA(0)=\lambda(0)P.\tag{13.2}$$

而且,$\mathscr{N}$中的特征向量$x(0)$满足

$$Px(0)=x(0). \tag{13.3}$$

于是,将 P 作用于(13.1),然后利用(13.2)和(13.3),可得

$$P\dot{A}(0)Px(0)+\lambda(0)P\dot{x}(0)=\dot{\lambda}(0)x(0)+\lambda(0)P\dot{x}(0),$$

亦即

$$P\dot{A}(0)Px(0)=\dot{\lambda}(0)x(0). \tag{13.4}$$

因 $A(t)$ 自伴,故 $\dot{A}(0)$ 自伴;又因 P 自伴,故 $P\dot{A}(0)P$ 自伴.显然, $P\dot{A}(0)P$ 映射 $\mathscr{N}$ 到 $\mathscr{N}$;关系式(13.4)是说,$\dot{\lambda}(0)$ 必须是 $P\dot{A}(0)P$ 在 $\mathscr{N}$ 上的一个特征值,$x(0)$ 是一个相应的特征向量.

3.13.2 定理 设 $A(t)=[a_{ij}(t)]$ 是自伴矩阵,其每个元素 a_{ij} 是实变量 t 的可微函数.假定在 $t=0$, $A(0)$ 有 $k>1$ 重特征值 λ_0.用 $\mathscr{N}$ 表示 $A(0)$ 属于 λ_0 的特征向量生成的 k 维空间,P 表示映射至 $\mathscr{N}$ 上的正交投影.假定 $\mathscr{N}$ 到 $\mathscr{N}$ 的自伴映射 $P\dot{A}(0)P$ 有 k 个不同的特征值

$$\mu_1,\cdots,\mu_k,$$

并用

$$y^{(1)},\cdots,y^{(k)}$$

表示相应的规范化的特征向量,则对足够小的 t, $A(t)$ 有 k 个特征值

$$\lambda_1(t),\cdots,\lambda_k(t)$$

近乎于 λ_0,具有如下性质:

(1) 每个 $\lambda_i(t)$ 可微地依赖于 t,而且随 $t\to 0$ 而趋于 λ_0.

(2) 当 $t\neq 0$ 时, $\lambda_1(t),\cdots,\lambda_k(t)$ 是不同的.

(3) $A(t)$ 属于 $\lambda_i(t)$ 的特征向量 $x^{(i)}(t)$,

$$A(t)x^{(i)}(t)=\lambda_i(t)x^{(i)}(t), \tag{13.5}$$

可以如此规范化,使得 $x^{(i)}(t)$ 随 $t\to 0$ 而趋于 $y^{(i)}$.

证 对于足够小的 t, $A(t)$ 的特征多项式和 $A(0)$ 的特征多项式

相差无几.根据假设,后者有 k 重根 λ_0;推出前者必有 k 个根随 $t \to 0$ 而趋于 λ_0.这些根是 $A(t)$ 的特征值 $\lambda_1(t),\cdots,\lambda_k(t)$,依 **3.2.3**,相应的特征向量

$$x^{(1)}(t),\cdots,x^{(k)}(t)$$

可经选取形成正交组.

下面先证明:随 $t \to 0$,每个规范化的特征向量 $x^{(i)}(t)$ 与特征向量空间 $\mathscr{N}$ 的距离趋于零;利用正交投影 P,可以表述为

$$\lim_{t\to 0}\left\|(I-P)x^{(i)}(t)\right\|=0,\quad i=1,\cdots,k. \tag{13.6}$$

利用 $t \to 0$ 时

$$A(t)\to A(0),\quad \lambda_i(t)\to\lambda_0,$$

又因 $\left\|x^{(i)}(t)\right\|=1$,从(13.5)推出

$$A(0)x^{(i)}(t)=\lambda_0 x^{(i)}(t)+\varepsilon(t). \tag{13.7}$$

注意,这里及以下,$\varepsilon(t)$ 是一个因不同场合而不同的笼统的记号,仅表示随 $t \to 0$ 而趋于零的向量或纯量.因 $\mathscr{N}$ 由 $A(0)$ 的特征向量组成,而且 P 投影任何向量至 $\mathscr{N}$ 上,

$$A(0)Px^{(i)}(t)=\lambda_0 Px^{(i)}(t). \tag{13.8}$$

从(13.7)减去(13.8),得

$$A(0)(I-P)x^{(i)}(t)=\lambda_0(I-P)x^{(i)}(t)+\varepsilon(t), \tag{13.9}$$

至此,可以采用反证法.

假定(13.6)不成立,则存在正数 d 和 $t \to 0$ 的一个序列,使得

$$\left\|(I-P)x^{(i)}(t)\right\|>d,$$

由于有限维空间中闭单位球是紧的;所以在所说的 $t \to 0$ 的序列中存在子序列,使得 $(I-P)x^{(i)}(t)$ 趋于一个非零极限 $x^{(0)}$.从(13.9)推出 $x^{(0)}$ 满足

$$A(0)x^{(0)}=\lambda_0 x^{(0)}, \tag{13.10}$$

于是 $x^{(0)}\in\mathscr{N}$.然而另一方面,每个向量 $(I-P)x^{(i)}(t)$ 正交于 $\mathscr{N}$,

其极限 $x^{(0)}$ 必正交于 $\mathscr{N}$;这与 $x^{(0)} \in \mathscr{N}$ 矛盾.故(13.6)成立.

现在进行 $x^{(i)}(t)$ 的连续性和 $\lambda_i(t)$ 的可微性的证明.从(13.5)减去(13.8)并且除以 t ;然后利用通常的 Leibniz 重排法,得

$$\frac{A(t)-A(0)}{t}x^{(i)}(t)+A(0)\frac{x^{(i)}(t)-Px^{(i)}(t)}{t}$$
$$=\frac{\lambda_i(t)-\lambda_i(0)}{t}x^{(i)}(t)+\lambda_i(0)\frac{x^{(i)}(t)-Px^{(i)}(t)}{t},$$

将 P 作用于等式;利用关系式(13.2),两边第二项相等.删去后得

$$P\frac{A(t)-A(0)}{t}x^{(i)}(t)=\frac{\lambda_i(t)-\lambda_i(0)}{t}Px^{(i)}(t). \quad (13.11)$$

因 A 已假定可微,

$$\frac{A(t)-A(0)}{t}=\dot{A}(0)+\varepsilon(t),$$

而且根据(13.6),

$$x^{(i)}(t)=Px^{(i)}(t)+\varepsilon(t),$$

将它们代入(13.11),并利用 $P^2=P$,有

$$P\dot{A}(0)PPx^{(i)}(t)=\frac{\lambda_i(t)-\lambda_i(o)}{t}Px^{(i)}(t)+\varepsilon(t). \quad (13.12)$$

按照假设, $\mathscr{N}$ 上自伴映射 $P\dot{A}(0)P$ 有 k 个不同特征值 $\mu_1,\cdots,\mu_k$,相应的规范化的特征向量是 $y^{(1)},\cdots,y^{(k)}$:

$$P\dot{A}(0)Py^{(j)}=\mu_j y^{(j)}, \quad j=1,\cdots k.$$

将 $Px^{(i)}$ 表示成这些特征向量的线性组合

$$Px^{(i)}(t)=\sum_{j=1}^{k}\alpha_{ij}(t)y^{(j)}, \quad (13.13)$$

代入(13.12),得

$$\sum_{j=1}^{k}\alpha_{ij}(t)\left(\mu_j-\frac{\lambda_i(t)-\lambda_i(0)}{t}\right)y^{(j)}=\varepsilon(t). \quad (13.14)$$

因为 $y^{(1)},\cdots,y^{(k)}$ 构成 $\mathscr{N}$ 的正交基,由(13.13)有

$$\left\|Px^{(i)}(t)\right\|^2=\sum_{j=1}^{k}\left|\alpha_{ij}(t)\right|^2.$$

再利用(13.6),以及规范化 $\left\|x^{(i)}(t)\right\|=1$,得

$$\sum_{j=1}^{k}\left|\alpha_{ij}(t)\right|^2=1-\varepsilon(t). \tag{13.15}$$

类似地,从(13.14)可得

$$\sum_{j=1}^{k}\left|\mu_j-\frac{\lambda_i(t)-\lambda_i(0)}{t}\right|^2\left|\alpha_{ij}(t)\right|^2=\varepsilon(t), \tag{13.16}$$

联合(13.15)和(13.16),推出对每个足够小的 t,在

$$\alpha_{i1}(t),\cdots,\alpha_{ik}(t)$$

中有一个绝对值为 $1-\varepsilon(t)$,不妨设其为 $\alpha_{ii}(t)$,因为如果不然,只须重排一下 $y^{(1)},\cdots,y^{(k)}$ 及相应的 $\mu_1,\cdots,\mu_k$.于是有

$$\left.\begin{array}{c}\left|\alpha_{ii}(t)\right|=1-\varepsilon(t);\\ \left|\alpha_{ij}(t)\right|\le\varepsilon(t),\forall j\ne i;\\ \left|\mu_i-\dfrac{\lambda_i(t)-\lambda_i(0)}{t}\right|\le\varepsilon(t).\end{array}\right\} \tag{13.17}$$

因为 $(\lambda_i(t)-\lambda_i(0))/t$ 当 $t\ne 0$ 时是 t 的连续函数,因此当 t 足够小时(13.17)中 μ_i 的指标 i 与 t 无关.

特征向量的规范化 $\left\|x^{(i)}(t)\right\|=1$ 仍保留一个绝对值为1的未定因子;选取此因子使得不仅 $\left|\alpha_{ii}(t)\right|$ 而且 $\alpha_{ii}(t)$ 本身近乎1:

$$\alpha_{ii}(t)=1-\varepsilon(t). \tag{13.18}$$

这样,联合(13.6),(13.13)和(13.17),推出

$$\left\|x^{(i)}(t)-y^{(i)}\right\|\le\varepsilon(t). \tag{13.19}$$

由于 $x^{(1)}(t),\cdots,x^{(k)}(t)$ 本身是正交组,显然两个正交单位向量不能同时与同一个向量 $y^{(i)}$ 之差的范数足够小,这就是说,不同的特征向量 $x^{(i)}(t)$ 近乎于不同的向量 $y^{(i)}$.

(13.19)表明,随 $t\to 0$,经适当规范化的 $x^{(i)}(t)$ 趋向 $y^{(i)}$.而(13.17)表明 $\lambda_i(t)$ 在 $t=0$ 可微,而且 $\dot{\lambda}_i(0)=\mu_i$;由此可知,当 t 足够小但不为零时,$A(t)$ 的近乎 λ_0 的特征值都是单根. □

3.14 稳定矩阵

3.14.1 引言 考虑一阶线性常系数常微分方程组

$$\frac{\mathrm{d}x}{\mathrm{d}t}=A\left(x(t)-\hat{x}\right), \tag{14.1}$$

其中 $A\in\mathbb{R}^{n\times n}$; $x(t),\hat{x}\in\mathbb{R}^n$.

显然,若在某时间 $\hat{t}$ 有

$$x(\hat{t})=\hat{x},$$

则 $x(t)$ 在 $t=\hat{t}$ 时停止变化.因此,$\hat{x}$ 称为方程组(14.1)的**平衡**.如果 A 是非奇异的,那么 $x(t)$ 仅当其达到平衡 $\hat{x}$ 时才停止变化.

关于这种方程组主要的问题是:

(1) 怎样给定初始值 $x(0)$ 能使 $t\to\infty$ 时 $x(t)\to\hat{x}$.

(2) 更重要地,在什么条件下,对于任意选取的 $x(0)$,都能使 $t\to\infty$ 时 $x(t)\to\hat{x}$.

如果对任意选取的初始值 $x(0)$,当 $t\to\infty$ 时均有 $x(t)\to\hat{x}$,则称 $\hat{x}$ 为方程组(14.1)的**整体稳定平衡**.

可以证明:平衡 $\hat{x}$ 整体稳定的充分必要条件是 A 的所有特征值有负实部.

更一般地,考虑一阶非线性常微分方程组

$$\frac{\mathrm{d}x}{\mathrm{d}t}=f\left(x(t)-\hat{x}\right), \tag{14.2}$$

其中

$$f:\mathbb{R}^n\to\mathbb{R}^n,\quad f=\left(f_1,\cdots,f_n\right)^{\mathrm{T}},$$

假定 $f_1,\cdots,f_n$ 有连续偏导数,而且 $f(x)=0$ 当且仅当 $x=0$.

仍然,$x(t)$ 停止变化当且仅当其达到平衡 $\hat{x}$;而且自然提出此平衡的稳定性问题.现在,存在多种可能的稳定性概念,它们不是全都能单用矩阵理论提出.

能单用矩阵理论提出的概念之一是**局部稳定性**.它与整体稳定性不同,问题的提法是:

对于选取的每个充分接近 $\hat{x}$ 的初始值 $x(0)$,当 $t\to\infty$ 时,$x(t)$ 是否收敛到 $\hat{x}$,也就是说,如果从平衡给以小扰动,系统是否收敛到平衡?

"充分接近"和"小"的说法可以加以精确化.

利用 Taylor 定理,方程组(14.2)可以近似地替换成线性常微分方程组

$$\frac{\mathrm{d}x}{\mathrm{d}t}=J_f\left(x(t)-\hat{x}\right),$$

而且在 $\hat{x}$ 的某个邻域内,不会改变原始方程定性的动态性质.这里

$$J_f\equiv\left[\frac{\partial f_i}{\partial x_j}(0)\right]$$

是取值于原点的 f 的 Jacobi 矩阵.

如果对在该邻域内的所有初始值 $x(0)$,当 $t\to\infty$ 时均有 $x(t)\to\hat{x}$,则称 $\hat{x}$ 为**局部稳定平衡**.

因此,平衡 $\hat{x}$ 局部稳定的充分必要条件是 J_f 的所有特征值有负实部.

3.14.2 定义 设 $A \in \mathbb{C}^{n\times n}$,用

$$i_+(A), i_-(A), i_0(A)$$

分别表示 A 按代数重数计算的具有正实部、负实部、零实部的特征值的个数,

$$i_+(A)+i_-(A)+i_0(A)=n, \tag{14.3}$$

三元组或说行向量

$$i(A) \equiv (i_+(A), i_-(A), i_0(A)) \tag{14.4}$$

称为 A 的**惯性**(inertia).

这一定义是 **3.1.2** 中自伴矩阵惯性概念的推广.

容易证明(留作练习)

$$\mathrm{i}(-A)^{\mathrm{T}} = \begin{bmatrix} 0 & 1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 1 \end{bmatrix} \mathrm{i}(A)^{\mathrm{T}}. \tag{14.5}$$

3.14.3 定义 设 $A \in \mathbb{C}^{n\times n}$,如果

$$i(A)=(0,n,0), \text{ 或者说 } i_-(A)=n,$$

则称 A 为**稳定矩阵**(stable matrix).如果

$$i(A)=(n,0,0), \text{ 或者说 } i_+(A)=n,$$

则称 A 为**正稳定矩阵**(positively stable matrix).

显然, $A \in \mathbb{C}^{n\times n}$ 稳定的充分必要条件是 $-A$ 正稳定.

因此,关于正稳定矩阵的任何定理通过合适的插入负号便可转述成关于稳定矩阵的定理.鉴于如此,纯粹出于数学上的方便,往往着重讨论正稳定矩阵的性质.

3.14.4 引理 设 $A\in\mathbb{C}^{n\times n}$ 是正稳定矩阵,则下列矩阵都是正稳定的:

(1) $\alpha A+\beta I,\forall\alpha\geq 0,\beta\geq 0,\alpha+\beta>0$.

(2) A^{-1}.

(3) A^{*}.

(4) A^{T}.

证明非常容易,留作练习.

3.14.5 引理 设 $A\in\mathbb{R}^{n\times n}$ 是正稳定矩阵,则

$$\det A^{k}>0,\quad k=1,2,\cdots \tag{14.6}$$

证 因 A 是实的,任何复特征值必共轭成对出现,其积作为正因子提供给行列式;又因 A 正稳定,任何实特征值都是正的,它们也是作为正因子提供给行列式.所以 $\det A>0$,并因此

$$\det A^{k}=(\det A)^{k}>0,\quad k=1,2,\cdots$$ □

3.14.6 Lyapunov 定理 设 $A\in\mathbb{C}^{n\times n}$.

(1) A 为正稳定矩阵的充分必要条件是存在正定矩阵 $G\in\mathbb{C}^{n\times n}$,使得

$$H\equiv GA+A^{*}G \tag{14.7}$$

是正定的.

(2) 若存在自伴矩阵 $G\in\mathbb{C}^{n\times n}$ 和正定矩阵 $H\in\mathbb{C}^{n\times n}$,成立

$$GA+A^{*}G=H, \tag{14.8}$$

则 A 为正稳定矩阵的充分必要条件是 G 为正定矩阵.

证 设 $S\in\mathbb{C}^{n\times n}$ 是任一非奇异矩阵,对(14.7)应用 * 相合变换,得到

$$S^{*}HS\equiv S^{*}GSS^{-1}AS+S^{*}A^{*}\left(S^{*}\right)^{-1}S^{*}GS.$$

简写成

$$\hat{H} \equiv \hat{G}\hat{A} + \hat{A}^*\hat{G}, \tag{14.9}$$

其中

$$\hat{H} = S^*HS, \quad \hat{G} = S^*GS, \quad \hat{A} = S^{-1}AS, \tag{14.10}$$

$\hat{H}$ 与 H 为 * 相合,$\hat{G}$ 与 G 为 * 相合,$\hat{A}$ 与 A 为相似.注意到 * 相合保自伴矩阵的惯性,相似性保特征值,而在本定理讨论中 H 和 G 总是自伴的.因此,可以假定 H,G,A 三个矩阵处于通过同时对 H 和 G 应用 * 相合变换及对 A 应用相似变换而能获得的任何一种特殊形式.

现在给出(1)的证明.假定 A 正稳定,$\hat{A}$ 是 A 的修正 Jordan 标准形,其任何非零的上次主对角元素全等于 $\varepsilon > 0$,此 ε 选取为小于 A 的每个特征值的实部;见[HJ1985]的推论(3.1.13).于是,若取 $\hat{G} = I_n$,则从(14.9)得

$$\hat{H} = \hat{A} + \hat{A}^*$$

是自伴矩阵,其对角元素是 $\hat{A}$ 的特征值的实部的2倍,全是正的,而且严格对角优势.因此,$\hat{H}$ 和 $\hat{G}$ 是正定的.必要性得证.

再证充分性.假定 G 和(14.7)确定的 H 都是正定矩阵.令

$$B \equiv GA,$$

则 $B + B^* = H$ 是正定的;对这两个等式同时左乘和右乘 G 的平方根 $G^{1/2}$ 的逆矩阵 $G^{-1/2}$,得

$$G^{1/2}AG^{-1/2} = G^{-1/2}BG^{-1/2}$$

和正定的

$$G^{-1/2}BG^{-1/2} + \left(G^{-1/2}BG^{-1/2}\right)^* = G^{-1/2}HG^{-1/2},$$

因此有

$$G^{1/2}AG^{-1/2} + \left(G^{1/2}AG^{-1/2}\right)^* = G^{-1/2}HG^{-1/2}.$$

这表明相似于 A 的 $G^{1/2}AG^{-1/2}$ 有正定的自伴部分,依 **3.12.2**,其特征值也就是 A 的特征值有正实部.

对于(2)的证明,充分性已蕴涵在(1)中,故只须证明必要性.

设 $A, G_1, H_1 \in \mathbb{C}^{n\times n}$，$A$ 是正稳定的，G_1 是自伴的，H_1 是正定的，而且成立

$$G_1 A + A^* G_1 = H_1,$$

要证明的是 G_1 必正定.

用反证法.假定 G_1 不是正定的，根据(1)，必存在正定矩阵 $G_2 \in \mathbb{C}^{n\times n}$，使得

$$G_2 A + A^* G_2 \equiv H_2$$

也是正定的.令

$$B_1 \equiv G_1 A, \quad B_2 \equiv G_2 A,$$

则有

$$B_i = \frac{B_i + B_i^*}{2} + \frac{B_i - B_i^*}{2} = \frac{1}{2}(H_i + S_i), \quad i = 1,2$$

其中

$$S_i \equiv B_i - B_i^*, \quad i = 1,2$$

是斜自伴的.

考虑自伴矩阵函数

$$G(\alpha) \equiv \alpha G_1 + (1-\alpha) G_2, \quad \alpha \in [0,1],$$

$G(0) = G_2$ 的所有特征值是正的，而 $G(1) = G_1$ 至少有一个特征值不是正的.

因为 $G(\alpha)$ 的特征值总是实的，而且连续地依赖于 α，所以存在 $\alpha_0 \in (0,1]$，使得 $G(\alpha_0)$ 至少有一个特征值是零，即 $G(\alpha_0)$ 为奇异的.

注意到

$$\begin{aligned} G(\alpha_0) A &= \alpha_0 B_1 + (1-\alpha_0) B_2 \\ &= \frac{1}{2}\alpha_0 H_1 + \frac{1}{2}(1-\alpha_0) H_2 \\ &\quad + \frac{1}{2}\alpha_0 S_1 + \frac{1}{2}(1-\alpha_0) S_2, \end{aligned}$$

有正定的自伴部分

$$\frac{G(\alpha_0)A+(G(\alpha_0)A)^*}{2}=\frac{1}{2}\alpha_0H_1+\frac{1}{2}(1-\alpha_0)H_2,$$

依 **3.12.2**,$G(\alpha_0)A$ 是非奇异的.然而,这与 $G(\alpha_0)$ 的奇异性相矛盾.因此,起初 G_1 非正定的假设不成立. □

通常,方程(14.8)称为 **Lyapunov 方程**,使 $GA+A^*G$ 成为正定矩阵的矩阵 G 统称为 Lyapunov 方程的 **Lyapunov 解**.

正定矩阵是正稳定矩阵的特殊情形.Lyapunov 定理作为正稳定矩阵的基本定理,表明了正定矩阵与所有正稳定矩阵的密切关系,提炼出以能够得到的(14.8)的右端来检验 $A\in\mathbb{C}^{n\times n}$ 的稳定性的可行步骤.

关于方程组(14.8),在 **7.1** 中有进一步的讨论.

可以想象:自伴矩阵是对实数的一种自然的矩阵模拟,正定矩阵是对正实数的一种自然的矩阵模拟,正稳定矩阵是对具有正实部的复数的一种自然的矩阵模拟.

3.14.7 定理 设 $A\in\mathbb{C}^{n\times n}$ 是正稳定矩阵.

(1) 对每个 $H\in\mathbb{C}^{n\times n}$,(14.8)有唯一解 G.

(2) 若 H 是自伴的,则相应的解 G 必是自伴的.

(3) 若 H 是正定的,则相应的解 G 必是正定的.

证 (1) 见 **7.1.20.**

(2) G 是(14.8)的解且 H 是自伴的,于是

$$H=H^*=(GA+A^*G)^*=G^*A+A^*G^*,$$

由此并根据(1)的解的唯一性结论,推出 $G^*=G$.

(3) 从 **3.14.6** 的(2)推出. □

3.14.8 推论 设 $A\in\mathbb{C}^{n\times n}$,$A$ 正稳定的充分必要条件是存在正定矩

阵 G 满足方程

$$GA + A^*G = I. \qquad \square \tag{14.11}$$

这是常见的 Lyapunov 定理的普通形式.

3.14.9 定义 设 $A \in \mathbb{C}^{n\times n}$,对于 $k=1,\cdots,n$,用

$$E_k(A)$$

表示 A 的 $\binom{n}{k}$ 个 k 阶主子式之和.

换句话说,$\pm E_k(A)$ 是 A 的特征多项式 $\det(\lambda I - A)$ 中 λ^{n-k} 的系数.特别

$$E_1(A) = \operatorname{tr}A, \quad E_n(A) = \det A$$

对于 $A \in \mathbb{R}^{n\times n}$,有另一个重要的稳定性判别准则,它着眼于 A 的主子式之和或特征多项式.

3.14.10 定义 设 $A \in \mathbb{R}^{n\times n}$,相伴 A 的 **Routh-Hurwitz 矩阵**是指如下 $\Omega(A) \in \mathbb{R}^{n\times n}$:

$$\Omega(A) \equiv [\omega_{ij}]$$

$$= \begin{bmatrix} E_1(A) & E_3(A) & E_5(A) & \cdots & \cdots & \cdots & \cdots & 0 \\ 1 & E_2(A) & E_4(A) & \cdots & \cdots & \cdots & \cdots & 0 \\ 0 & E_1(A) & E_3(A) & & & & & \vdots \\ 0 & 1 & E_2(A) & \ddots & & & & \vdots \\ 0 & 0 & E_1(A) & & \ddots & & & \vdots \\ \vdots & \vdots & 1 & & & \ddots & & \vdots \\ \vdots & \vdots & \vdots & & & & \ddots & 0 \\ 0 & 0 & 0 & \cdots & \cdots & \cdots & \cdots & E_n(A) \end{bmatrix} \tag{14.12}$$

其中记号$E_i(A)(i=1,\cdots,n)$如 **3.14.9** 定义.

$\Omega(A)$的结构是:对于$i=1,\cdots,n$,有

(1) 对角元素$\omega_{ii}=E_i(A)$.

(2) 在第i列中:

ω_{ii}之上,

$$\omega_{i-k,i}=E_{i+k}(A),k=1,\cdots,i-1,$$

这里,规定当$j>n$时$E_j(A)=0$;

ω_{ii}之下,

$$\omega_{i+k,i}=E_{i-k}(A),k=1,\cdots,n-i,$$

这里,规定$E_0(A)=1$,当$j<0$时$E_j(A)=0$.

3.14.11 Routh-Hurwitz 稳定判别准则 矩阵$A\in\mathbb{R}^{n\times n}$正稳定的充分必要条件是 Routh-Hurwitz 矩阵$\Omega(A)$的前主子式是正的.

证明从略.

描述一种新的证明技巧:利用$E_i(A)$写出A的特征多项式,再构造一个以如此多项式作为其特征多项式的简单形式的矩阵,诸如友矩阵;然后应用 Lyapunov 定理(需经许多代数计算)得出 Routh-Hurwitz 条件.

关于 Routh-Hurwitz 稳定判别准则有两点值得指出:

(1) 在有些场合,它是确定$A\in\mathbb{R}^{n\times n}$惯性的更一般而复杂条件的一种特殊情况.

(2) 它可以等价地陈述为检测多项式的“正稳定性”,即多项式的所有零点是否分布在复平面的右半平面.

3.14.12 定义 矩阵$A\in\mathbb{C}^{n\times n}$称为**半正稳定**的,如果

$$i_-(A)=0,$$

亦即 A 的每个特征值有非负实部.

3.14.13 引理 设 $A\in\mathbb{C}^{n\times n}$,若存在正定的 $G\in\mathbb{C}^{n\times n}$ 和半正定的 $H\in\mathbb{C}^{n\times n}$ 成立(14.8),则 A 是半正稳定的.

证 定义

$$S\equiv GA-A^{*}G. \tag{14.13}$$

显然,若 G 自伴,则 S 斜自伴.从(14.8)和(14.13)得

$$2GA=H+S. \tag{14.14}$$

现设 $\lambda\in\lambda(A)$ 且 $Ax=\lambda x$, $x\neq 0$,于是

$$2GAx=2\lambda Gx,$$

并从(14.14)有

$$(H+S)x=2\lambda Gx. \tag{14.15}$$

左乘 x^{*} 得

$$2\lambda x^{*}Gx=x^{*}Hx+x^{*}Sx. \tag{14.16}$$

对(14.16)取实部,注意到

$$x^{*}Gx>0,\ \ x^{*}Hx\geq 0,$$

以及依 **3.4.2** 知 $x^{*}Sx$ 是纯虚数,得

$$2(\operatorname{Re}\lambda)x^{*}Gx=x^{*}Hx, \tag{14.17}$$

由此推出 $\operatorname{Re}\lambda\geq 0$. □

3.14.14 定理 设 $A\in\mathbb{C}^{n\times n}$,并设存在正定的 $G\in\mathbb{C}^{n\times n}$ 和半正定的 $H\in\mathbb{C}^{n\times n}$ 成立(14.8),而且 S 如(14.13)定义,则 A 为正稳定的充分必要条件是在 H 的零空间 $\mathcal{N}_H$ 中没有 $G^{-1}S$ 的特征向量.

证 假设有 $x\neq 0$ 是 $G^{-1}S$ 的特征向量,而且 $x\in\mathcal{N}_H$,则 $Hx=0$,从而对某个 $\lambda\in\mathbb{C}$成立

$$\lambda Gx=Sx=(S+H)x=2GAx,$$

推出

$$Ax=\frac{1}{2}\lambda x.$$

由于$\frac{1}{2}\lambda$是A的特征值,根据(14.17),

$$\operatorname{Re}\lambda = 2\operatorname{Re}\left(\frac{1}{2}\lambda\right) = \frac{x^* Hx}{x^* Gx} = 0,$$

因此A不是正稳定的.这表明条件“A正稳定”蕴涵“$\mathscr{N}_H$中没有$G^{-1}S$的特征向量”.

另一方面,依 **3.14.13**, A的所有特征值有非负实部.如若A不是正稳定的,则存在$\lambda \in \lambda(A)$, $\operatorname{Re}\lambda = 0$.于是,当$x \neq 0$是$A$的相应于$\lambda$的特征向量时,从(14.17)推出

$$x^* Hx = 0,$$

而如此情况因H是半正定的,仅当$Hx = 0$时才能发生.将$Hx = 0$代入(14.15),得

$$Sx = 2\lambda Gx,$$

因此,x是$G^{-1}S$的特征向量,$x \in \mathscr{N}_H$.这表明条件“$\mathscr{N}_H$中没有$G^{-1}S$的特征向量”蕴涵“A是正稳定的”. □

3.14.15 定义 设$A \in \mathbb{C}^{n\times n}$, $B \in \mathbb{C}^{n\times m}$,**矩阵对**$(A,B)$称为**可控制的**(controllable),如果矩阵

$$\mathscr{C} = \begin{bmatrix} B & AB & A^2B & \cdots & A^{n-1}B \end{bmatrix} \tag{14.18}$$

的秩

$$\operatorname{rank}\mathscr{C} = n, \tag{14.19}$$

$\mathscr{C}$称为**可控制矩阵**(controllability matrix).

可控制性的概念出现在线性控制微分方程组理论中.

设$A \in \mathbb{R}^{n\times n}$和$B \in \mathbb{R}^{n\times m}$是给定的矩阵,向量值函数$u$称为**控制**,它是由控制变量组成的,向量

$$u(t) = \left(u_1(t), \cdots, u_m(t)\right)^{\mathrm{T}} \in \mathbb{R}^m$$

而且,对任意给定的连续的u,以及对每一给定的向量$x^{(0)} \in \mathbb{R}^n$,初值问题

$$\frac{\mathrm{d}x}{\mathrm{d}t}=Ax(t)+Bu(t),\quad x(0)=x^{(0)} \tag{14.20}$$

有唯一解.

可控制性问题如下:对于每对 $\hat{x},x^{(0)}\in\mathbb{R}^n$,是否存在某个时刻 $\hat{t}<\infty$ 及控制 u 的某种选取,使得初值问题(14.20)的解满足 $x(\hat{t})=\hat{x}$? 答案是:肯定存在的充分必要条件是矩阵对 (A,B) 是可控制的.

在应用上,可控制的意义是可能操纵控制变量,使系统从任一初始状态启动,实现达到任一所希望的状态.例如,通过适当地控制火箭发动机,将太空船从围绕地球的轨道转移到指定的围绕月球的轨道,或者"软着陆"至月球.

已经证明:**3.14.14** 中有关 $G^{-1}S$ 的条件的等价说法是矩阵对 (A^*,H) 为可控制的.因此,若正定的 $G\in\mathbb{C}^{n\times n}$ 和半正定的 $H\in\mathbb{C}^{n\times n}$ 满足(14.8),则 $A\in\mathbb{C}^{n\times n}$ 为正稳定的充分必要条件是矩阵对 (A^*,H) 为可控制的.

3.14.16 定理 设 $A\in\mathbb{C}^{n\times n}$,存在自伴矩阵 $G\in\mathbb{C}^{n\times n}$ 和正定矩阵 $H\in\mathbb{C}^{n\times n}$ 成立(14.8)的充分必要条件是

$$i_0(A)=0.$$

而且,满足如此条件时,成立

$$i(A)=i(G).$$

证明从略.可参见[HJ1991]中的(2.4.10).

3.14.17 定理 $A\in\mathbb{C}^{n\times n}$ 正稳定的充分必要条件是对每个非零的 $x\in\mathbb{C}^n$ 存在正定矩阵 $G\in\mathbb{C}^{n\times n}$ 使得

$$\mathrm{Re}(x^*GAx)>0. \tag{14.21}$$

证 设 $x\in\mathbb{C}^n$ 是 A 的相应于 $\lambda\in\lambda(A)$ 的特征向量,并设 $G\in\mathbb{C}^{n\times n}$ 是任一正定矩阵.则

$$\operatorname{Re}(x^*GAx)=\operatorname{Re}(x^*G(\lambda x))=\operatorname{Re}(x^*Gx\lambda)=(x^*Gx)\operatorname{Re}\lambda,$$

因 $x^*Gx>0$,故 $\operatorname{Re}(x^*GAx)$ 与 $\operatorname{Re}\lambda$ 同号.这样,充分性已然得证.

再证必要性.由于 A 是正稳定的,应用 **3.14.6**,存在正定的 G 使得由(14.7)定义的 H 也是正定的.因此,对于任何非零的 $x\in\mathbb{C}^n$,

$$\operatorname{Re}(x^*GAx)=\frac{1}{2}(x^*GAx+x^*A^*Gx)=\frac{1}{2}x^*Hx>0. \qquad \square$$

Lyapunov 定理保证:如果寻找到单一的正定矩阵 G,使得二次型 x^*GAx 对所有非零的 $x\in\mathbb{C}^n$ 有正的实部,那么 A 是正稳定的. **3.14.17** 表明,这个全局要求(对所有 x 存在一个 G)可以放松到非常局部的要求(对每个 x 存在一个 G).

4 非负矩阵

4.1 基本概念和基本性质

4.1.1 定义 设 $A=[a_{ij}]\in\mathbb{R}^{m\times n}$,如果

$$a_{ij}\geq 0,\quad \forall i=1,\cdots,m,\ j=1,\cdots,n, \tag{1.1}$$

即 A 的所有元素是非负的,则称 A 为**非负矩阵**(nonnegative matrix),记作 $A\geq 0$.

如果(1.1)中成立严格不等号,即 A 的所有元素是正数,则称 A 为**正矩阵**(positive matrix).记作 $A>0$.

设 $A,B\in\mathbb{R}^{m\times n}$,如果成立 $A-B\geq 0$,则记作 $A\geq B$;如果成立 $A-B>0$,则记作 $A>B$.

注意:正(非负)矩阵与正(非负)自伴映射使用同样的记号,不要把概念混淆了.

对于任意的 $A=[a_{ij}]\in\mathbb{C}^{m\times n}$,引进记号

$$|A|=\left[\,|a_{ij}|\,\right], \tag{1.2}$$

即表示以 a_{ij} 之模 $|a_{ij}|$ 为元素的矩阵.

以上定义自然适用于向量.n 维行向量是 $1\times n$ 矩阵,n 维列向量是 $n\times 1$ 矩阵.

4.1.2 定理 设 $A,B,C,D\in\mathbb{C}^{m\times n}$,$\alpha,\beta\in\mathbb{C}$,则

(1) $|A| \geq 0$；$|A| = 0$ 当且仅当 $A = 0$.

(2) $|\alpha A| = |\alpha||A|$.

(3) $|A + B| \leq |A| + |B|$.

(4) 若 $A \geq 0$ 且 $A \neq 0$，则当 m 或 n 大于1时不一定成立 $A > 0$.

(5) 若 $A \geq 0, B \geq 0$，且 $\alpha, \beta \geq 0$，则 $\alpha A + \beta B \geq 0$.

(6) 若 $A \geq B, C \geq D$，则 $A + C \geq B + D$.

(7) 若 $A \geq B, B \geq C$，则 $A \geq C$.

证明留作练习.

4.1.3 定理 设 $A, B, C, D \in \mathbb{C}^{n \times n}, x \in \mathbb{C}^n$，则

(1) $|Ax| \leq |A||x|$.

(2) $|AB| \leq |A||B|$.

(3) $|A^m| \leq |A|^m, m = 1, 2, \cdots$.

(4) 若 $0 \leq A \leq B, 0 \leq C \leq D$，则 $0 \leq AC \leq BD$.

(5) 若 $0 \leq A \leq B$，则 $0 \leq A^m \leq B^m, m = 1, 2, \cdots$.

(6) 若 $A \geq 0$，则 $A^m \geq 0$；若 $A > 0$，则 $A^m > 0, m = 1, 2, \cdots$.

(7) 若 $A > 0, x \geq 0$ 且 $x \neq 0$，则 $Ax > 0$.

(8) 若 $A \geq 0, x > 0$ 且 $Ax = 0$，则 $A = 0$.

(9) 若 $|A| \leq |B|$，则 $\|A\|_F \leq \|B\|_F$.

(10) $\| |A| \|_F = \|A\|_F$.

证明留作练习.

显然，论断(9)和(10)对任何绝对范数皆成立，Frobenius 范数仅是一个例子.

4.1.4 定理 设 $A, B \in \mathbb{C}^{n \times n}, |A| \leq B$，则

$$\rho(A) \leq \rho(|A|) \leq \rho(B). \tag{1.3}$$

证 依 **4.1.3** 的(3)和(5),

$$\left|A^m\right| \le |A|^m \le B^m,\ \ m = 1,2,\cdots$$

由此并依 **4.1.3** 的(9)和(10),有

$$\left\|A^m\right\|_F \le \left\||A|^m\right\|_F \le \left\|B^m\right\|_F,\ \ m = 1,2,\cdots$$

从而

$$\left\|A^m\right\|_F^{1/m} \le \left\||A|^m\right\|_F^{1/m} \le \left\|B^m\right\|_F^{1/m},\ \ m = 1,2,\cdots$$

现在让 $m \to \infty$,而且应用 **2.1.26**,即得(1.3). □

4.1.5 推论 设 $A, B \in \mathbb{R}^{n\times n}$,且 $0 \le A \le B$,则

$$\rho(A) \le \rho(B).$$ □

4.1.6 推论 设 $A \in \mathbb{R}^{n\times n}$, $A \ge 0$, $\tilde{A}$ 是 A 的任一主子矩阵,则

$$\rho(\tilde{A}) \le \rho(A).$$

特别

$$\max_{1\le i\le n} a_{ii} \le \rho(A). \tag{1.4}$$

证 设 $\tilde{A}$ 是 A 的 $r\times r$ 主子矩阵, $1 \le r \le n$.用 $\hat{A}$ 表示将 A 中除 $\tilde{A}$ 的元素所在位置以外的元素置为零而得的 $n\times n$ 矩阵.于是有

$$\rho(\tilde{A}) = \rho(\hat{A}), \qquad 0 \le \hat{A} \le A.$$

因此,依 **4.1.5**,

$$\rho(\tilde{A}) = \rho(\hat{A}) \le \rho(A).$$ □

4.1.7 引理 设 $A \in \mathbb{R}^{n\times n}$, $A \ge 0$.

(1) 若 A 的各行之和相等,则 $\rho(A) = \|A\|_\infty$.

(2) 若 A 的各列之和相等,则 $\rho(A) = \|A\|_1$.

证 依 **2.1.20**,对于任何矩阵范数$\|\bullet\|$, $\rho(A)\le\|A\|$.

如果 A 的各行之和相等,那么 $x=(1,\cdots,1)^{\mathrm T}$ 是 A 的相应特征值 $\|A\|_\infty$ 的特征向量,因此 $\rho(A)=\|A\|_\infty$.

如果 A 的各列之和相等,那么现在 $x=(1,\cdots,1)^{\mathrm T}$ 是 $A^{\mathrm T}$ 的相应特征值$\|A^{\mathrm T}\|_\infty$ 的特征向量,因此

$$\rho(A)=\rho(A^{\mathrm T})=\|A^{\mathrm T}\|_\infty=\|A\|_1. \qquad \square$$

4.1.8 定理 设 $A=[a_{ij}]\in\mathbb{R}^{n\times n}$, $A\ge 0$,则成立

$$\min_{1\le i\le n}\sum_{j=1}^n a_{ij}\le\rho(A)\le\max_{1\le i\le n}\sum_{j=1}^n a_{ij} \tag{1.5}$$

和

$$\min_{1\le j\le n}\sum_{i=1}^n a_{ij}\le\rho(A)\le\max_{1\le j\le n}\sum_{i=1}^n a_{ij}. \tag{1.6}$$

证 记

$$\alpha\equiv\min_{1\le i\le n}\sum_{j=1}^n a_{ij}.$$

构造 $B=[b_{ij}]\in\mathbb{R}^{n\times n}$,使得

$$A\ge B\ge 0;\quad \sum_{j=1}^n b_{ij}=\alpha,\quad i=1,\cdots,n.$$

比如,若 $\alpha=0$,则取 $B=0$;若 $\alpha>0$,则取

$$b_{ij}=\alpha\frac{a_{ij}}{\sum\limits_{j=1}^n a_{ij}},\quad 1\le i,j\le n.$$

依 **4.1.7**, $\rho(B)=\alpha$;再依 **4.1.5**,

$$\rho(B)\le\rho(A),$$

此即(1.5)中行和形式的下界.按相仿方式容易确立(1.5)中行和形式

的上界.

将行和形式的界应用于 A^{T},即得(1.6)中列和形式的界. □

4.1.9 推论 设 $A=[a_{ij}]\in\mathbb{R}^{n\times n}$, $A\geq 0$ 且

$$\sum_{j=1}^{n}a_{ij}>0,\quad i=1,\cdots n,$$

则 $\rho(A)>0$.

特别,若 $A>0$,或者 A 是不可约非负矩阵,则 $\rho(A)>0$. □

4.1.10 定理 设 $A=[a_{ij}]\in\mathbb{R}^{n\times n}$, $A\geq 0$,则对任何

$$x=(x_1,\cdots,x_n)^{\mathrm{T}}\in\mathbb{R}^n,\quad x>0,$$

成立

$$\min_{1\leq i\leq n}\frac{1}{x_i}\sum_{j=1}^{n}a_{ij}x_j\leq\rho(A)\leq\max_{1\leq i\leq n}\frac{1}{x_i}\sum_{j=1}^{n}a_{ij}x_j \tag{1.7}$$

和

$$\min_{1\leq j\leq n}x_j\sum_{i=1}^{n}\frac{a_{ij}}{x_i}\leq\rho(A)\leq\max_{1\leq j\leq n}x_j\sum_{i=1}^{n}\frac{a_{ij}}{x_i}. \tag{1.8}$$

证 取

$$S=\mathrm{diag}(x_1,\cdots,x_n),$$

从 $A\geq 0$ 和 $x>0$ 推出

$$S^{-1}AS=\left[a_{ij}\frac{x_j}{x_i}\right]\geq 0.$$

将 **4.1.8** 应用于 $S^{-1}AS$,并注意到

$$\rho(S^{-1}AS)=\rho(A),$$

即得(1.7)和(1.8). □

4.1.11 推论 设 $A=[a_{ij}]\in\mathbb{R}^{n\times n}$, $A\geq 0$; $x=(x_1,\cdots,x_n)^{\mathrm{T}}\in\mathbb{R}^n$,且

$x > 0$.如果

$$\alpha, \beta \in \mathbb{R},\ \alpha, \beta \geq 0,\ \alpha x \leq Ax \leq \beta x,$$

那么

$$\alpha \leq \rho(A) \leq \beta.$$

而且,若 $\alpha x < Ax$,则 $\alpha < \rho(A)$;若 $Ax < \beta x$,则 $\rho(A) < \beta$.

证 设 $\alpha x \leq Ax$,则有

$$\alpha \leq \min_{1 \leq i \leq n} \frac{1}{x_i} \sum_{j=1}^{n} a_{ij} x_j,$$

依 **4.1.10**,得 $\alpha \leq \rho(A)$.而当 $\alpha x < Ax$ 时,必存在

$$\alpha' > \alpha, \quad \alpha' x \leq Ax,$$

从而

$$\rho(A) \geq \alpha' > \alpha,$$

上界的推导相仿. □

4.1.12 推论 设 $A \in \mathbb{R}^{n \times n}$, $A \geq 0$.如果 A 有正特征向量,那么对应的特征值必是 $\rho(A)$;也就是说,如果 $Ax = \lambda x$ 且 $x > 0$,那么 $\lambda = \rho(A)$.

证 设 $x > 0$ 且 $Ax = \lambda x$,则 $\lambda \geq 0$,从而可将 $Ax = \lambda x$ 写成

$$\lambda x \leq Ax \leq \lambda x,$$

于是,依 **4.1.11**,

$$\lambda \leq \rho(A) \leq \lambda. \qquad \square$$

4.1.13 推论 设 $A = [a_{ij}] \in \mathbb{R}^{n \times n}$, $A \geq 0$.如果 A 有正特征向量,那么

$$\rho(A) = \max_{x>0} \min_{1 \leq i \leq n} \frac{1}{x_i} \sum_{j=1}^{n} a_{ij} x_j = \min_{x>0} \max_{1 \leq i \leq n} \frac{1}{x_i} \sum_{j=1}^{n} a_{ij} x_j. \tag{1.9}$$

证 设

$$\hat{x}=(\hat{x}_1,\cdots,\hat{x}_n)^{\mathrm{T}}>0$$

是 A 的特征向量.依 **4.1.12**,成立 $A\hat{x}=\rho(A)\hat{x}$.由此,显然有

$$\min_{1\le i\le n}\frac{1}{\hat{x}_i}\sum_{j=1}^{n}a_{ij}\hat{x}_j=\rho(A)=\max_{1\le i\le n}\frac{1}{\hat{x}_i}\sum_{j=1}^{n}a_{ij}\hat{x}_j .$$

再注意到依 **4.1.10** 有

$$\min_{1\le i\le n}\frac{1}{\hat{x}_i}\sum_{j=1}^{n}a_{ij}\hat{x}_j\le\max_{x>0}\min_{1\le i\le n}\frac{1}{x_i}\sum_{j=1}^{n}a_{ij}x_j\le\rho(A)$$

和

$$\rho(A)\le\min_{x>0}\max_{1\le i\le n}\frac{1}{x_i}\sum_{j=1}^{n}a_{ij}x_j\le\max_{1\le i\le n}\frac{1}{\hat{x}_i}\sum_{j=1}^{n}a_{ij}\hat{x}_j ,$$

即得(1.9). □

4.1.14 推论 设 $A=[a_{ij}]\in\mathbb{R}^{n\times n}$,$A\ge 0$.如果 A 有正特征向量 $x=(x_1,\cdots,x_n)^{\mathrm{T}}$,那么对于 $m=1,2,\cdots$ 和 $i=1,\cdots,n$,成立

$$\sum_{j=1}^{n}a_{ij}^{(m)}\le\left(\frac{\max\limits_{1\le k\le n}x_k}{\min\limits_{1\le k\le n}x_k}\right)\rho(A)^m \tag{1.10}$$

和

$$\left(\frac{\min\limits_{1\le k\le n}x_k}{\max\limits_{1\le k\le n}x_k}\right)\rho(A)^m\le\sum_{j=1}^{n}a_{ij}^{(m)} , \tag{1.11}$$

其中 $A^m\equiv\left[a_{ij}^{(m)}\right]$.因此,如果 A 有正特征向量且 $\rho(A)>0$,那么对于 $m=1,2,\cdots$,矩阵 $\left(\rho(A)^{-1}A\right)^m$ 的元素一致有界.

证 依 **4.1.12**, $Ax=\rho(A)x$.从而

$$A^m x=\rho(A)^m x .$$

再注意到 $A^m\ge 0$,有

$$\rho(A)^m \max_{1\le k\le n} x_k \ge \rho(A)^m x_i = (A^m x)_i \equiv \sum_{j=1}^n a_{ij}^{(m)} x_j$$

$$\ge \left(\min_{1\le k\le n} x_k\right)\sum_{j=1}^n a_{ij}^{(m)}, \quad i = 1,\cdots,n,$$

此式因 $x > 0$ 而等价于(1.10).类似地,有

$$\rho(A)^m \min_{1\le k\le n} x_k \le \rho(A)^m x_i = (A^m x)_i \equiv \sum_{j=1}^n a_{ij}^{(m)} x_j$$

$$\le \left(\max_{1\le k\le n} x_k\right)\sum_{j=1}^n a_{ij}^{(m)}, \quad i = 1,\cdots,n.$$

此式因 $x > 0$ 而等价于(1.11). □

4.2 正矩阵和不可约非负矩阵

4.2.1 定理 设 $A \in \mathbb{R}^{n\times n}$, $A \ge 0$,则 A 不可约的充分必要条件是

$$(I + A)^{n-1} > 0.$$

更一般地,设 $A \in \mathbb{C}^{n\times n}$,则 A 不可约的充分必要条件是

$$(I + |A|)^{n-1} > 0.$$

证 先证必要性.条件 $(I + A)^{n-1} > 0$ 等价于对任一非零向量 $x \ge 0$,成立

$$(I + A)^{n-1} x > 0.$$

因此,可以转而考虑从 $x^{(0)} = x$ 出发的非负向量序列

$$x^{(k+1)} = (I + A)x^{(k)}, \quad k = 0,1,\cdots,n-2.$$

只须证明当 $x^{(k)}$ 有零分量时,$x^{(k+1)}$ 的非零分量必比 $x^{(k)}$ 多,便可从 $x^{(0)}$ 至少有1个正分量递归地推出 $x^{(n-1)} = (I + A)^{n-1} x > 0$.

显然，$x^{(k+1)}$ 的非零分量不会比 $x^{(k)}$ 少.如果 $x^{(k+1)}$ 和 $x^{(k)}$ 的非零分量正好都是 m 个，$1 \le m < n$，那么从 $x^{(k)}$ 到 $x^{(k+1)}$ 的递推关系容易看出存在置换矩阵 P，使得

$$Px^{(k+1)} = \begin{bmatrix} y \\ 0 \end{bmatrix}, \quad y > 0; \quad Px^{(k)} = \begin{bmatrix} z \\ 0 \end{bmatrix}, \quad z > 0,$$

向量 y 和 z 有 m 个分量.于是

$$\begin{bmatrix} y \\ 0 \end{bmatrix} = Px^{(k+1)} = Px^{(k)} + PAP^{\mathrm{T}}Px^{(k)} = \begin{bmatrix} z \\ 0 \end{bmatrix} + \begin{bmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{bmatrix} \begin{bmatrix} z \\ 0 \end{bmatrix},$$

其中 A_{11} 是 $m \times m$ 矩阵.由此，$A_{21}z = 0$，并从 $A_{21} \ge 0$ 及 $z > 0$ 推出 $A_{21} = 0$，但这与 A 不可约相矛盾.

再证充分性.假设在 $(I + A)^{n-1} > 0$ 的条件下非负矩阵 A 可约，即存在置换矩阵 P，使得

$$PAP^{\mathrm{T}} = \begin{bmatrix} A_{11} & A_{12} \\ 0 & A_{22} \end{bmatrix}, \quad A_{11} \in \mathbb{R}^{m \times m}, \quad m < n,$$

那么

$$P(I + A)^{n-1} P^{\mathrm{T}} = \begin{bmatrix} I + A_{11} & A_{12} \\ 0 & I + A_{22} \end{bmatrix}^{n-1} = \begin{bmatrix} (I + A_{11})^{n-1} & * \\ 0 & (I + A_{22})^{n-1} \end{bmatrix},$$

这与 $(I + A)^{n-1} > 0$ 矛盾. □

正矩阵是不可约矩阵.

非负矩阵则有可约和不可约两类.特别，关于不可约非负矩阵已有不少结论.

4.2.2 引理 设 $A \in \mathbb{R}^{n \times n}$，$A \ge 0$，且存在

$$x \in \mathbb{R}^n, x \ge 0, x \ne 0; \quad \alpha \in \mathbb{R},$$

满足 $Ax > \alpha x$,则

$$\rho(A) > \alpha .$$

证 $\alpha < 0$ 时不用证明,只须考虑 $\alpha \geq 0$.取 $\varepsilon > 0$,使其满足

$$Ax \geq (\alpha + \varepsilon)x .$$

并令 $B = A/(\alpha + \varepsilon)$,则成立

$$B^k x \geq B^{k-1} x \geq \cdots \geq Bx \geq x, \quad k = 1,2,\cdots$$

由此及 $x \geq 0, x \neq 0$,推出 $k \to \infty$ 时 B^k 不趋于零.于是,依 **2.1.23**, $\rho(B) \geq 1$,即得

$$\rho(A) \geq \alpha + \varepsilon > \alpha . \qquad \square$$

4.2.3 定理 设 $A = [a_{ij}] \in \mathbb{R}^{n \times n}$, $A > 0$,则 $\rho(A)$ 是 A 的具有如下性质的**优势特征值**:

(1) $\rho(A) > 0$,它是 A 的特征值,而且对应特征向量可取作正的,即存在 $u \in \mathbb{R}^n$,

$$Au = \rho(A)u, \quad u > 0 . \tag{2.1}$$

(2) $\rho(A)$ 的几何重数是1.

(3) $|\lambda| < \rho(A), \forall \lambda \in \lambda(A)$ 且 $\lambda \neq \rho(A)$.

因此,对于任意取定的向量范数 $\|\bullet\|$,存在唯一的向量

$$u = (u_1, \cdots, u_n)^{\mathrm{T}} > 0, \quad \|u\| = 1$$

满足(2.1).

证 由于对于任何矩阵 A 均存在的特征值 $\lambda \in \mathbb{C}$ 及对应特征向量 $v = (v_1, \cdots, v_n)^{\mathrm{T}} \in \mathbb{C}^n, v \neq 0$,满足

$$Av = \lambda v, \quad |\lambda| = \rho(A). \tag{2.2}$$

由此并注意到现在有 $A > 0$,

$$|\lambda v_i| = \left| \sum_{j=1}^{n} a_{ij} v_j \right| \leq \sum_{j=1}^{n} a_{ij} |v_j|, \quad j = 1, \cdots, n . \tag{2.3}$$

利用数学归纳法容易证明:在 $A > 0$ 的条件下,(2.3)中各式等号

成立的充分必要条件为存在仅依赖于向量v的单位复数η, $|\eta|=1$, 使得

$$\eta v_j \geq 0, \quad j=1,\cdots,n. \tag{2.4}$$

证明留作练习.[提示:条件(2.4)等价于存在θ, $0\leq\theta<2\pi$,使得

$$v_j = r_j e^{i\theta}, \quad r_j \geq 0, \quad j=1,\cdots,n,$$

即$v_1,\cdots,v_n$位于复平面从原点出发的射线$z=re^{i\theta}$上.]

因此,如若(2.4)不成立,那么(2.3)中取严格不等号,即有

$$|\lambda||v| < A|v|, \tag{2.5}$$

于是,依 **4.2.2**, $\rho(A)>|\lambda|$,这与假设$|\lambda|=\rho(A)$矛盾.这样,(2.4)是成立的.取$u=\eta v$,则$u\geq 0$,(2.2)等价于

$$Au=\lambda u, \quad |\lambda|=\rho(A).$$

由此得知$\lambda>0$,从而

$$\lambda=|\lambda|=\rho(A),$$

而且进一步有$u>0$.(1)得证.

以上推证已蕴涵(3)的结论.

现在证明(2).

结论(1)的一个明显推论是:若z是任一相应$\rho(A)$的特征向量,则$|z|>0$.

假定$\rho(A)$的几何重数大于1,那么至少存在线性无关的特征向量$u=(u_1,\cdots,u_n)^{\mathrm{T}}$和$v=(v_1,\cdots,v_n)^{\mathrm{T}}$,

$$Au=\rho(A)u, \quad |u|>0; \quad Av=\rho(A)v, \quad |v|>0,$$

取

$$z=v-\frac{v_1}{u_1}u,$$

则z也是相应$\rho(A)$的特征向量,应有$|z|>0$,但其第一个分量为零,矛盾.因此$\rho(A)$的几何重数为1. □

4.2.4 Perron-Frobenius 定理 设 $A \in \mathbb{R}^{n\times n}$ 是不可约的非负矩阵,则

(1) A 存在一个正的实特征值等于其谱半径 $\rho(A)$.

(2) $\rho(A)$对应的特征向量可取作正的.

(3) $\rho(A)$是 A 的单特征值.

(4) 不存在相应 A 的其它特征值的非负特征向量.

证 先证(1)和(2).根据谱半径的定义,A 存在特征值 λ 和相应的特征向量 $x \in \mathbb{C}^n$,满足

$$Ax = \lambda x, \quad |\lambda| = \rho(A),$$

因此,利用 A 的非负性,有

$$\rho(A)|x| = |\lambda||x| \le A|x|.$$

由此递推,得

$$(\rho(A))^k |x| \le A^k |x|, \quad k = 1,2,\cdots \tag{2.6}$$

于是成立

$$(1+\rho(A))^{n-1}|x| \le (I+A)^{n-1}|x|. \tag{2.7}$$

依 **4.2.1**,

$$(I+A)^{n-1} > 0, \quad (I+A^T)^{n-1} > 0.$$

将 **4.2.3** 应用于 $(I+A^{\mathrm{T}})^{n-1}$,得知其有特征值 $\rho((I+A)^{n-1})$及存在相应特征向量 $y>0$;从而利用转置可得

$$y^{\mathrm{T}}(I+A)^{n-1} = \rho((I+A)^{n-1})y^{\mathrm{T}}. \tag{2.8}$$

用 y^{T} 左乘(2.7),然后应用(2.8)并且约去因子 $y^T|x|>0$,有

$$(1+\rho(A))^{n-1} \le \rho((I+A)^{n-1}). \tag{2.9}$$

另一方面,$(I+A)^{n-1}>0$ 的特征值是

$$(1+\mu)^{n-1}, \ \forall \mu \in \lambda(A).$$

依 **4.2.3**,存在 $\hat{\mu} \in \lambda(A)$,使得

$$\rho((I+A)^{n-1}) = (1+\hat{\mu})^{n-1}.$$

将其代入(2.9),可得

$$1+\rho(A) \le |1+\hat{\mu}| \le 1+|\hat{\mu}| \le 1+\rho(A),$$

由此推出

$$\hat{\mu}=|\hat{\mu}|=\rho(A).$$

因而(2.9)成立等号,并因此(2.7)成立等号,即

$$(I+A)^{n-1}|x|=(1+\rho(A))^{n-1}|x|.$$

根据这一等式,连同$(I+A)^{n-1}>0$及$|x|\ne 0$,推出$|x|>0$;再连同(2.6),有

$$A|x|=\rho(A)|x|.$$

这表明:$\rho(A)>0$是A的特征值,$|x|>0$是相应的特征向量.

现在证(3).由以上推证已知,属于$\rho(A)$的任一特征向量$\boldsymbol{u}$必有$|\boldsymbol{u}|>0$.因此,仿照**4.2.3**可证$\rho(A)$的几何重数为1.另外,对A和A^{T}应用(2)的结论,属于$\rho(A)$的左、右特征向量$\boldsymbol{u}$,$\boldsymbol{v}$可取作正的,因而$\boldsymbol{v}^{\mathrm{T}}\boldsymbol{u}>0$.因此,依**2.1.30**,$\rho(A)$是$A$的单特征值.

最后证(4).反证法.假定存在非零的非负向量z满足

$$Az=\lambda z, \quad \lambda \ne \rho(A). \tag{2.10}$$

从(1)和(2)的推证已知,存在$u>0$,使得

$$u^{\mathrm{T}}A=\rho(A)u^{\mathrm{T}}.$$

由此,并利用(2.10),有

$$\rho(A)u^{\mathrm{T}}z=\lambda u^{\mathrm{T}}z.$$

注意到$\boldsymbol{u}^{\mathrm{T}}z>0$,推出$\lambda=\rho(A)$,这与$\lambda\ne\rho(A)$的假定矛盾.因此(4)的结论成立. □

4.2.3 表明正矩阵A只有一个模等于$\rho(A)$的特征值,这就是$\rho(A)$本身.**4.2.4** 则表明不可约的非负矩阵A除了$\rho(A)$本身是特征值外,还可以存在若干模等于$\rho(A)$的特征值.

4.2.5 引理 设$A\in\mathbb{C}^{n\times n}$,$\lambda\in\mathbb{C}$,$x,y\in\mathbb{C}^n$,使得

$$Ax = \lambda x, \quad A^{\mathrm{T}} y = \lambda y, \quad x^{\mathrm{T}} y = 1 \tag{2.11}$$

定义 $L \equiv xy^{\mathrm{T}}$,则

(1) $Lx = x$, $y^{\mathrm{T}} L = y^{\mathrm{T}}$.

(2) $L^m = L$, $m = 1,2,\cdots$.

(3) $A^m L = LA^m = \lambda^m L$, $m = 1,2,\cdots$.

(4) $L(A - \lambda L) = 0$.

(5) $(A - \lambda L)^m = A^m - \lambda^m L$, $m = 1,2,\cdots$.

(6) $A - \lambda L$ 的每个非零特征值也是 A 的特征值.

如果附加条件

$$\lambda \neq 0, \lambda \text{ 是 } A \text{ 的几何重数为1的特征值} \tag{2.12}$$

那么

(7) λ 不是 $A - \lambda L$ 的特征值,亦即 $\lambda I - (A - \lambda L)$ 是可逆的.

最后,如果假定

$$|\lambda| = \rho(A) > 0, \lambda \text{ 是 } A \text{ 的唯一模为 } \rho(A) \text{ 的特征值}, \tag{2.13}$$

而且 A 的特征值排列成

$$|\lambda_1| \leq \cdots \leq |\lambda_{n-1}| < |\lambda_n| = |\lambda| = \rho(A),$$

那么

(8) $\rho(A - \lambda L) \leq |\lambda_{n-1}| < \rho(A)$.

(9) $(\lambda^{-1} A)^m = L + (\lambda^{-1} A - L)^m \to L$,当 $m \to \infty$.

(10) 对于使得

$$(|\lambda_{n-1}| / \rho(A)) < r < 1$$

的每个 r,存在常数 $c = c(r, A)$,使得

$$\left\|(\lambda^{-1} A)^m - L\right\|_\infty < cr^m, \quad m = 1,2,\cdots.$$

证 由条件(2.11)直接推出(1),(2),(3)和(4);而且利用归纳法即

可推出(5).注意,$x^{\mathrm{T}}y=1$蕴涵x和y都是非零向量.

设$\mu\neq 0$是$A-\lambda L$的特征值,$w\neq 0$是相应μ的特征向量,

$$(A-\lambda L)w=\mu w.$$

利用(4),有

$$\mu Lw=L(A-\lambda L)w=0w=0,$$

因此$Lw=0$.从而

$$Aw=(A-\lambda L)w=\mu w,$$

这表明μ也是A的特征值.(6)得证.

现在考虑条件(2.12).如果$\mu=\lambda$是$A-\lambda L$的特征值,w是相应的特征向量,那么因$\lambda\neq 0$,根据(6)的论证,w也是A的相应λ的特征向量.由于λ是A的几何重数为1的特征值,必有$w=\alpha x$,α是某一非零常数.因此,利用(1),有

$$\mu w=\lambda w=(A-\lambda L)w=(A-\lambda L)\alpha x=\alpha\lambda x-\lambda\alpha x=0,$$

但这是不可能的,因为$\lambda\neq 0$,$w\neq 0$.矛盾.(7)得证.

从(6)得知,对于$\rho(A-\lambda L)$来说,或者

$$\rho(A-\lambda L)=|\lambda_k|,\ \lambda_k\text{ 是 }A\text{ 的某个特征值},$$

或者

$$\rho(A-\lambda L)=0,$$

鉴于A的特征值按模递增排序,而且$|\lambda_n|=|\lambda|=\rho(A)$.因此,在条件(2.13)下,利用(6)和(7),即得(8).

联合(8)和(5).从(8)有

$$\rho(\lambda^{-1}A-L)=\frac{\rho(A-\lambda L)}{\rho(A)}\leq\frac{|\lambda_{n-1}|}{\rho(A)}<1,$$

由此并利用(5),得

$$(\lambda^{-1}A-L)^m=(\lambda^{-1}A)^m-L\to 0,\text{当 }m\to\infty,$$

这等价于(9).

至于(10)的收敛速度的结论,只要将**2.1.25**应用于矩阵

$$\lambda^{-1}A-L,$$

并选取 ε 使得

$$\rho\left(\lambda^{-1}A-L\right)+\varepsilon\le\left(\left|\lambda_{n-1}\right|/\rho(A)\right)+\varepsilon<r<1,$$

即可得证. □

4.2.6 定理 设 $A\in\mathbb{R}^{n\times n}, A>0$,则

$$\lim_{m\to\infty}\left(\rho(A)^{-1}A\right)^m=L, \tag{2.14}$$

其中 $L\equiv xy^{\mathrm{T}}, x>0, y>0$,而且

$$Ax=\rho(A)x,\ \ A^{\mathrm{T}}y=\rho(A)y,\ \ x^{\mathrm{T}}y=1.$$

证 因为

$$A>0, A^{\mathrm{T}}>0, \rho(A)=\rho\left(A^{\mathrm{T}}\right),$$

依 **4.2.3**,存在 $x>0, z>0$,使得

$$Ax=\rho(A)x,\quad A^{\mathrm{T}}z=\rho(A)z.$$

取

$$y\equiv\left(x^{\mathrm{T}}z\right)^{-1}z,$$

这里 $x^Tz>0$ 是实数,从而有

$$x^{\mathrm{T}}y=x^{\mathrm{T}}\left(x^{\mathrm{T}}z\right)^{-1}z=\left(x^{\mathrm{T}}z\right)^{-1}\left(x^{\mathrm{T}}z\right)=1.$$

现在,对于 $\lambda=\rho(A)$, x 和 y 来说,满足 **4.2.5** 的条件(2.11),(2.12)和(2.13).因此,从 **4.2.5** 的(9)得(2.14). □

4.2.7 推论 设 $A\in\mathbb{R}^{n\times n}, A>0$,则

$$L\equiv\lim_{m\to\infty}\left(\rho(A)^{-1}A\right)^m$$

是秩为1的正矩阵.

4.2.8 定理(Ky Fan) 设 $A=[a_{ij}]\in\mathbb{C}^{n\times n}, B=[b_{ij}]\in\mathbb{R}^{n\times n}$,而且 $B\ge0, |A|\le B$,则

$$\lambda(A)\subset\bigcup_{i=1}^{n}\left\{z\in:\left|z-a_{ii}\right|\le\rho(B)-b_{ii}\right\}. \tag{2.15}$$

证 注意,**4.1.6** 保证

$$\rho(B)-b_{ii}\geq 0,\quad i=1,\cdots,n.$$

可以假定 $B>0$.因为若 B 的某些元素为零,则可转而考虑矩阵 $B_\varepsilon\equiv[b_{ij}+\varepsilon],\varepsilon>0;B_\varepsilon>|A|$,当 $\varepsilon\to 0$ 时

$$\rho(B_\varepsilon)-(b_{ii}+\varepsilon)\to\rho(B)-b_{ii}.$$

现在,$B>0$,依 **4.2.3**,存在向量 $x=(x_1,\cdots,x_n)^{\mathrm{T}}>0$,使得

$$Bx=\rho(B)x.$$

从而

$$\sum_{j=1,j\neq i}^{n}|a_{ij}|x_j\leq\sum_{j=1,j\neq i}^{n}b_{ij}x_j=\rho(B)x_i-b_{ii}x_i,\ i=1,\cdots,n.$$

于是,有

$$\frac{1}{x_i}\sum_{j=1,j\neq i}^{n}|a_{ij}|x_j\leq\rho(B)-b_{ii},\ i=1,\cdots,n.$$

至此,从 **1.6.25** 可直接推出(2.15). □

4.2.9 定理 设 $A\in\mathbb{R}^{n\times n},A\geq 0,A$ 有正的左特征向量.而且设存在 $x\in\mathbb{R}^n,x\geq 0,x\neq 0$,满足

$$Ax\geq\rho(A)x,$$

则

$$Ax=\rho(A)x.$$

证 设 $y>0$ 是 A 的左特征向量,$A^{\mathrm{T}}y=\rho(A)y$,则

$$y^{\mathrm{T}}\left(Ax-\rho(A)x\right)=\rho(A)y^{\mathrm{T}}x-\rho(A)y^{\mathrm{T}}x=0.$$

由此并注意到 $Ax-\rho(A)x\geq 0$,推出

$$Ax-\rho(A)x=0.$$ □

4.2.10 定理 设 $A,B\in\mathbb{C}^{n\times n},|A|\leq B,B$ 不可约,$\rho(B)=\rho(A)$,而且

$$\lambda=\mathrm{e}^{\mathrm{i}\varphi}\rho(A)$$

是 A 的特征值,则存在

$$\theta_1,\cdots,\theta_n \in \mathbb{R},$$

使得

$$A = e^{i\varphi} D B D^{-1}, \tag{2.16}$$

其中

$$D = \operatorname{diag}\left(e^{i\theta_1},\cdots,e^{i\theta_n}\right).$$

证 由假设

$$|\lambda| = \rho(A) = \rho(B),$$

且存在 $x = \left(x_1,\cdots,x_n\right)^{\mathrm{T}} \neq 0$,使得 $Ax = \lambda x$,有

$$\rho(B)|x| = |\lambda x| = |Ax| \leq |A||x| \leq Bx. \tag{2.17}$$

因为 $B \geq 0$ 且不可约,依 **4.2.4**, B 有正的左特征向量;于是,根据(2.17),依 **4.2.9**,得

$$B|x| = \rho(B)|x|. \tag{2.18}$$

并因而

$$|Ax| = |A||x| = B|x|. \tag{2.19}$$

这样,依 **4.4.2**,(2.18)蕴涵 $|x| > 0$,再依 **4.1.3** 的(8),即从(2.19)推出

$$|A| = B.$$

现在定义

$$\theta_k \in \mathbb{R}, \quad e^{i\theta_k} \equiv \frac{x_k}{|x_k|}, \quad k = 1,\cdots,n.$$

并令

$$D \equiv \operatorname{diag}\left(e^{i\theta_1},\cdots,e^{i\theta_n}\right),$$

则有

$$x = D|x|, \quad \lambda x = Ax = AD|x| = e^{i\varphi}\rho(B)D|x|,$$

因此

$$e^{-i\varphi} D^{-1} A D|x| = \rho(B)|x| = B|x|.$$

再注意到

$$|x|>0,\ \left|e^{-i\varphi}D^{-1}AD\right|=B,$$

推出 $e^{-i\varphi}D^{-1}AD=B$. □

4.3 循环矩阵和素矩阵

4.3.1 定义 设 $A\in\mathbb{R}^{n\times n}$ 是不可约的非负矩阵,且有 k 个模等于 $\rho(A)$ 的特征值.

如果 $k=1$,则称 A 为**素矩阵**(primitive matrix).

如果 $k>1$,则称 A 是**指数为 k 的循环矩阵**(cyclic matrix of index k)或**非素矩阵**(imprimitive matrix).

循环矩阵和素矩阵的概念是为讨论非负矩阵的谱而引入的.

4.3.2 定理 设 $A\in\mathbb{R}^{n\times n}$ 是指数为 k 的循环矩阵,也就是说,A 是 $n\times n$ 的不可约非负矩阵,而且有 k 个模等于 $\rho(A)$ 的特征值,则

(1) A 的模等于 $\rho(A)$ 的 k 个特征值是

$$\lambda_j=\rho(A)\mathrm{e}^{\mathrm{i}\frac{2j\pi}{k}},\quad j=0,1,\cdots,k-1. \tag{3.1}$$

几何上看,$\lambda_0,\lambda_1,\cdots,\lambda_{k-1}$ 均匀地分布在复平面的以原点为圆心、$\rho(A)$ 为半径的圆周上.

(2) A 的特征多项式 $p(\lambda)\equiv\det(\lambda I-A)$ 形如

$$p(\lambda)=\lambda^m\left(\lambda^k-\rho(A)^k\right)\left(\lambda^k-\delta_2\rho(A)^k\right)\cdots\left(\lambda^k-\delta_r\rho(A)^k\right), \tag{3.2}$$

其中 $rk+m=n$,当 $r>1$ 时

$$0<|\delta_i|<1,\ i=2,\cdots,r.$$

几何上看,除 k 个模等于 $\rho(A)$ 的特征值外,A 的其余非零特征值亦

可分组,每组均有 k 个模相等的特征值均匀地分布在复平面的圆心在原点、半径小于 $\rho(A)$ 的某圆周上.

(3) 存在 $n\times n$ 的置换矩阵 P,使得

$$PAP^{\mathrm{T}}=\begin{bmatrix} 0 & \cdots & \cdots & 0 & A_{1k} \\ A_{21} & 0 & & & 0 \\ 0 & A_{32} & 0 & & \vdots \\ \vdots & \ddots & \ddots & \ddots & \vdots \\ 0 & \cdots & 0 & A_{k,k-1} & 0 \end{bmatrix}, \qquad (3.3)$$

其中对角零子矩阵是方阵,(3.3)称为循环矩阵 A 的**标准形**.

此定理证明冗长,可参见[HJ1985]的 8.4.6 或[V1962]的 2.2.

例 (1) 设 $n\times n$ 非负矩阵

$$A=\begin{bmatrix} 0 & \cdots & \cdots & 0 & 1 \\ 1 & \ddots & & & 0 \\ 0 & 1 & \ddots & & \vdots \\ \vdots & \ddots & \ddots & \ddots & \vdots \\ 0 & \cdots & 0 & 1 & 0 \end{bmatrix},$$

可证(留作练习) A 不可约,其特征值是

$$\lambda_j=\mathrm{e}^{\mathrm{i}\frac{2j\pi}{n}}, \quad \rho(A)=\left|\lambda_j\right|=1, \quad j=1,\cdots,n.$$

因此 A 是指数为 n 的循环矩阵(标准形).这是简单的典型情形.

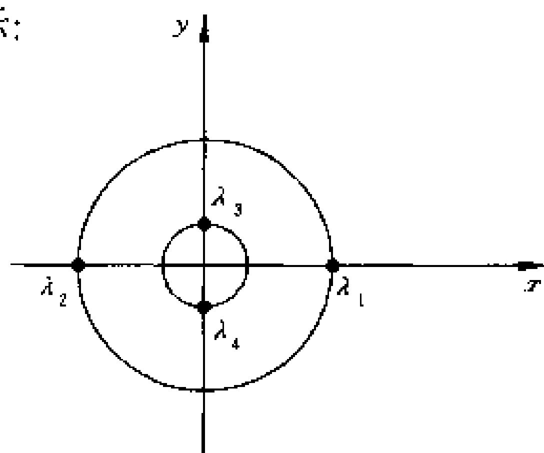
(2) 设非负矩阵

$$A=\begin{bmatrix} 0 & 0 & 1 & 0 \\ 0 & 0 & 1 & 1 \\ 0 & 1 & 0 & 0 \\ 1 & 1 & 0 & 0 \end{bmatrix},$$

容易证明 A 是不可约的.它的特征值是

$$\lambda_{1,2}=\pm\sqrt{\sqrt{2}+1},\quad \lambda_{3,4}=\pm\sqrt{\sqrt{2}-1},$$

其分布如下图所示:



因此, A 是 $k=2$ 的循环矩阵.而且

$$\rho(A)=\lambda_1=\sqrt{\sqrt{2}+1}.$$

A 可写成

$$A=\begin{bmatrix}0 & A_{12}\\ A_{21} & 0\end{bmatrix},\quad A_{12}=\begin{bmatrix}1 & 0\\ 1 & 1\end{bmatrix},\quad A_{21}=\begin{bmatrix}0 & 1\\ 1 & 1\end{bmatrix},$$

表明其本身已是标准形.

4.3.3 定理 设 $A=[a_{ij}]\in\mathbb{R}^{n\times n}$ 是指数为 $k\ (>1)$ 的循环矩阵,记 $A^m\equiv\left[a_{ij}^{(m)}\right]$,则对于不是 k 的整数倍的任何正整数 m ,成立

$$a_{ii}^{(m)}=0,\quad i=1,\cdots,n.$$

特别

$$a_{ii}\equiv a_{ii}^{(1)}=0,\ i=1,\cdots,n.$$

证 依 **4.3.2**,可选取 A 的一个模最大的特征值为

$$\lambda=\mathrm{e}^{\mathrm{i}\varphi}\rho(A),\quad \varphi=\frac{2\pi}{k}.$$

故当正整数 m 不是 k 的整数倍时, $\mathrm{e}^{\mathrm{i}m\varphi}$ 不是实数或正数.

依 **4.2.10**,取 $B = A$,有

$$A = e^{i\varphi} DAD^{-1},$$

从而

$$A^m \equiv \left[a_{ij}^{(m)}\right] = e^{im\varphi} DA^m D^{-1},$$

推出

$$a_{ii}^{(m)} = e^{im\varphi} a_{ii}^{(m)}, \quad i = 1, \cdots, n; \ m = 1, 2, \cdots$$

可见,只要 $e^{im\varphi}$ 不是实数或正数,便不可能有 $a_{ii}^{(m)} > 0$.因此,当 m 不是 k 的整数倍时,必有 $a_{ii}^{(m)} = 0$, $i = 1, \cdots, n$. □

4.3.4 定义 $A \in \mathbb{C}^{n\times n}$(不必非负或不可约)称为**指数为** k (>1) 的**弱循环矩阵**,如果存在 $n \times n$ 置换矩阵 P,使得 PAP^{T} 形如(3.3),其中零对角子矩阵是方阵.

4.3.5 定理 设 $A \in \mathbb{C}^{n\times n}$ 是指数为 k (>1) 的弱循环矩阵,则

$$p(\lambda) \equiv \det(\lambda I - A) = \lambda^m \prod_{j=1}^{r} \left(\lambda^k - \sigma_j^k\right). \tag{3.4}$$

证明从略.

4.3.6 引理 设 $A \in \mathbb{R}^{n\times n}$, $A > 0$,则 A 是素矩阵.

证 A 是非负的和不可约的;而且依 **4.2.4** 和 **4.2.3**, $\rho(A)$是单特征值, A 的任一不等于 $\rho(A)$的特征值 λ 必有 $|\lambda| < \rho(A)$,即只有 1 个模为 $\rho(A)$的特征值. □

4.3.7 引理 设 A 是素矩阵.则对所有正整数 m, A^m 也是素矩阵.

证 因为 $\rho(A)$是 A 的单特征值,而且是 A 的模为 $\rho(A)$的唯一特征值,所以 $(\rho(A))^m$ 是 A^m 的单特征值,而且是 A^m 的模为 $(\rho(A))^m$ 的唯一特征值.因此,对于 $m \geq 1$,由于有 $A^m \geq 0$,要证明 A^m 是素矩阵,只须再证明 A^m 是不可约的.

用反证法.假定对某一正整数 v, A^v 是可约的而且不妨设其形如

$$A^v = \begin{bmatrix} B & C \\ 0 & D \end{bmatrix}, \tag{3.5}$$

其中 B 和 D 是方子矩阵.设 $Ax = \rho(A)x$, $x > 0$,于是

$$A^v x = (\rho(A))^v x,$$

对应于 B 和 D,将 x 写成

$$x = \begin{bmatrix} x^{(1)} \\ x^{(2)} \end{bmatrix},$$

则从(3.5),

$$Dx^{(2)} = (\rho(A))^v x^{(2)},$$

表明 $(\rho(A))^v$ 是 D 的特征值.

另一方面,再考虑 A^{T},它也是非负的和不可约的,依 **4.2.4**,存在向量 $y > 0$ 使得 $A^{\mathrm{T}} y = \rho(A) y$.从(3.5),类似地可推出 $(\rho(A))^v$ 是 B^{T} 的特征值,从而也是 B 的特征值.

这样,$(\rho(A))^v$ 是 D 的特征值又是 B 的特征值,$\rho(A)$成为了 A 的多重特征值,矛盾.因此,对于所有 $m \geq 1$, A^m 是不可约的; A^m 是素矩阵. □

4.3.8 引理 设 $A = [a_{ij}] \in \mathbb{R}^{n \times n}$, $A \geq 0$, A 不可约,而且

$$a_{ii} > 0, \quad i = 1, \cdots, n,$$

则 $A^{n-1} > 0$.

证 显然,可以选取 $n \times n$ 的不可约非负矩阵 B,使得

$$A \geq \gamma(I + B), \quad \gamma > 0.$$

于是,依 **4.1.2**,

$$A^{n-1} \geq \gamma^{n-1}(I+B)^{n-1} > 0. \qquad \square$$

4.3.9 定理 设 $A=[a_{ij}]\in\mathbb{R}^{n\times n}$, $A\geq 0$,则存在某正整数 m 使得 $A^m>0$ 的充分必要条件是 A 为素矩阵.

证 若 $n\times n$ 矩阵 B 可约,则存在 $n\times n$ 置换矩阵 P,

$$PBP^{\mathrm{T}}=\begin{bmatrix} B_{11} & B_{12} \\ 0 & B_{22} \end{bmatrix},$$

其中 B_{11} 和 B_{22} 是方阵.于是,PBP^T 的幂形如

$$PB^mP^{\mathrm{T}}=\left(PBP^{\mathrm{T}}\right)^m=\begin{bmatrix} B_{11}^m & B_{12}^{(m)} \\ 0 & B_{22}^m \end{bmatrix},$$

这说明每个可约矩阵的幂也是可约的.

现在假设对某正整数 m, $A^m>0$,那么 A 是不可约的,而且依 **4.3.6**, A^m 是素矩阵.如果 A 不是素矩阵,则 A 是指数为某个 $k>1$ 的循环矩阵,亦即 A 存在 k 个模为 $\rho(A)$ 的特征值.因此, A^m 有 k 个模为 $(\rho(A))^m$ 的特征值,这与 A^m 是素矩阵矛盾.必要性得证.

反之,假设 A 是素矩阵,根据定义,A 是不可约的.依 **10.2.9** , A 存在由非零元素组成的从第1行某元素 a_{1,j_1} 到第1列某元素 $a_{i_1,1}$ 的长为 m_1 的闭路径;再依 **10.1.6**,这等价于 $A^{m_1}\equiv\left[a_{ij}^{(1)}\right]$ 有 $a_{11}^{(1)}>0$.注意到依 **4.3.7**, A^{m_1} 仍是素矩阵,必不可约,因而存在由非零元素组成的从第2行某元素 $a_{2,j_2}^{(1)}$ 到第2列某元素 $a_{i_2,2}^{(1)}$ 的长为 m_2 的闭路径,等价于 $\left(A^{m_1}\right)^{m_2}\equiv\left[a_{ij}^{(2)}\right]$ 有 $a_{22}^{(2)}>0$.依此类推,最后有 $A^{m_1m_2\cdots m_n}$ 是不可约的,非负的,其所有对角元素是正的.至此,依 **4.3.8** 得知,A 的某次幂是正的. $\square$

4.3.10 定理 设 $A\in\mathbb{C}^{n\times n}$ 是指数为 k (>1) 的弱循环矩阵,则对每个

整数 $j \geq 1$, A^{jk} 是完全可约的,也就是说,存在 $n \times n$ 置换矩阵 P,使得

$$PA^{jk}P^{\mathrm{T}} = \begin{bmatrix} C_1^j & 0 & \cdots & 0 \\ 0 & C_2^j & \ddots & \vdots \\ \vdots & \ddots & \ddots & 0 \\ 0 & \cdots & 0 & C_k^j \end{bmatrix}, \tag{3.6}$$

其中每个对角子矩阵 C_i 是方阵,而且

$$\rho(C_1) = \rho(C_2) = \cdots = \rho(C_k) = (\rho(A))^k. \tag{3.7}$$

进一步地,若 A 是非负、不可约和指数为 k 的循环矩阵,则每个子矩阵 C_i 是素矩阵.

证 直接计算(3.3)的 PAP^{T} 的幂,表明(3.6)的 $PA^{jk}P^{\mathrm{T}}$ 的表示形式对每个 $j \geq 1$ 成立,而且

$$C_i = A_{i,i-1} \cdots A_{21} A_{1k} A_{k,k-1} \cdots A_{i+1,i}, \quad i = 1, \cdots, k.$$

注意:

$$C_1 = A_{1k} A_{k,k-1} \cdots A_{21}, \quad C_k = A_{k,k-1} \cdots A_{21} A_{1k}.$$

由于从 $C_i x = \lambda x$ 推出

$$C_{i+1}(A_{i+1,i}x) = A_{i+1,i} C_i x = \lambda (A_{i+1,i} x),$$

表明所有 C_i 有相同的特征值,因此它们有相同的谱半径,成立(3.7).

余下的证明留作练习. □

4.3.11 定理 设 $A \in \mathbb{R}^{n \times n}$ 是素矩阵,则

$$\lim_{m \to \infty} \left(\rho(A)^{-1} A\right)^m = L > 0,$$

其中 $L \equiv xy^{\mathrm{T}}$, $x > 0$, $y > 0$,而且

$$Ax = \rho(A)x, \; A^{\mathrm{T}} y = \rho(A) y, \; x^{\mathrm{T}} y = 1.$$

同时,如果 λ_{n-1} 是 A 的特征值,使得

$$\frac{|\lambda_{n-1}|}{\rho(A)}<r<1;\quad |\lambda|\le|\lambda_{n-1}|,\quad \forall\lambda\in\lambda(A),\lambda\ne\rho(A),$$

则存在常数 $c=c(r,A)$,成立

$$\left\|\left(\rho(A)^{-1}A\right)^m-L\right\|_\infty\le cr^m,\quad m=1,2,\cdots$$

证 依 **4.2.5**,并仿照 **4.2.6** 的证明,即得结论. □

4.4 可约非负矩阵

4.4.1 定理 设 $A\in\mathbb{R}^{n\times n}$ 是非负矩阵,则存在 $n\times n$ 置换矩阵 P,使得

$$PAP^T=\begin{bmatrix}A_{11}&A_{12}&\cdots&A_{1k}\\0&A_{22}&\cdots&A_{2k}\\\vdots&\ddots&\ddots&\vdots\\0&\cdots&0&A_{kk}\end{bmatrix},\tag{4.1}$$

其中 $A_{ii}\ (i=1,\cdots,k)$ 是不可约非负方阵.

证 若 A 不可约,则结论已成立,

$$A=A_{11},\quad k=1.$$

现设 A 可约.依 **2.2.1**,存在 $n\times n$ 置换矩阵 P_1,使得

$$P_1AP_1^T=\begin{bmatrix}A_{11}^{(1)}&A_{12}^{(1)}\\0&A_{22}^{(1)}\end{bmatrix},\tag{4.2}$$

其中对角块 $A_{11}^{(1)}$ 和 $A_{22}^{(1)}$ 分别是 $r\times r$ 子矩阵和 $(n-r)\times(n-r)$ 子矩阵,$1\le r<n$.

于是,可以问 $A_{11}^{(1)}$ 和 $A_{22}^{(1)}$ 是否不可约,如若两者均不是,则分别存在 $r\times r$ 置换矩阵 P_{21} 和 $(n-r)\times(n-r)$ 置换矩阵 P_{22},使得

$$P_{21}A_{11}^{(1)}P_{21}^{\mathrm{T}}=\begin{bmatrix}A_{11}^{(2)} & A_{12}^{(2)}\\ 0 & A_{22}^{(2)}\end{bmatrix},\quad P_{22}A_{22}^{(1)}P_{22}^{\mathrm{T}}=\begin{bmatrix}A_{33}^{(2)} & A_{34}^{(2)}\\ 0 & A_{44}^{(2)}\end{bmatrix}.$$

因此

$$P_2P_1AP_1^{\mathrm{T}}P_2^{\mathrm{T}}=\begin{bmatrix}A_{11}^{(2)} & A_{12}^{(2)} & A_{13}^{(2)} & A_{14}^{(2)}\\ 0 & A_{22}^{(2)} & A_{23}^{(2)} & A_{24}^{(2)}\\ 0 & 0 & A_{33}^{(2)} & A_{34}^{(2)}\\ 0 & 0 & 0 & A_{44}^{(2)}\end{bmatrix},$$

其中

$$P_2=\begin{bmatrix}P_{21} & 0\\ 0 & P_{22}\end{bmatrix}.$$

自然,对凡已不可约的对角块即可停止变换.如此,最后可变换成(4.1)的形式. □

(4.1)称为**可约矩阵的标准形**.

4.4.2 定理 设 $A\in\mathbb{R}^{n\times n}$ 是非负矩阵,则

(1) $\rho(A)$是 A 的特征值,其对应特征向量可取作非负的.

(2) A 的特征值可分成若干组,每组中特征值的模都相等,而且均匀地分布在复平面的以原点为圆心的某个圆周上.

(3) 若 A 有一个正的特征向量 x,则 x 必是属于 $\rho(A)$的特征向量.

证 对于 A 不可约的情形,定理的结论可直接从 4.2.4 和 4.3.2 推出.

现在考虑 A 可约的情形.假设 A 的标准形是(4.1).如果(4.1)中某 A_{ii} 是1×1 零矩阵,则其单特征值为零.除如此情形外,由于任一子矩阵 A_{ii} 是不可约且非负的,因而 A_{ii} 有正特征值等于其谱半径$\rho(A_{ii})$.于是,A 显然有一非负的特征值等于$\rho(A)$.

注意,如若

$$\rho(A)=0,$$

则(4.1)中的每个 A_{ii} 是 1×1 零矩阵,即(4.1)的矩阵是严格上三角矩阵.

本定理的余下结论利用 **4.2.4** 和 **4.3.2** 及标准形(4.1)不难加以推证,可作练习. □

这一定理是 **4.2.4** 和 **4.3.2** 的延伸.

4.4.3 定义 设 $A\in\mathbb{R}^{n\times n}$, $A\geq 0$,则 $\rho(A)$ 是 A 的特征值,而且存在相应于 $\rho(A)$ 的右特征向量

$$x=(x_1,\cdots,x_n)^{\mathrm{T}}\geq 0,\quad \sum_{i=1}^{n}x_i=1;\quad Ax=\rho(A)x,$$

和左特征向量

$$y=(y_1,\cdots,y_n)^{\mathrm{T}}\geq 0,\quad \sum_{i=1}^{n}y_i=1;\quad y^{\mathrm{T}}A=\rho(A)y^{\mathrm{T}},$$

$\rho(A)$ 称为 A 的 **Perron 特征值**, x 和 y 分别称为 A 的**右 Perron(特征)向量和左 Perron(特征)向量**.

4.4.4 定理 设 $A\in\mathbb{R}^{n\times n}$, $A\geq 0$;并设存在

$$x\in\mathbb{R}^n,\ x\geq 0,\ x\neq 0;\quad \alpha\in\mathbb{R}$$

满足 $Ax\geq\alpha x$.则

$$\rho(A)\geq\alpha.$$

特别,当 $Ax>\alpha x$ 时

$$\rho(A)>\alpha.$$

证 设 $A=[a_{ij}]$, $\varepsilon\in\mathbb{R}$, $\varepsilon>0$,则

$$A(\varepsilon)\equiv[a_{ij}+\varepsilon]>0.$$

依 **4.2.3**, $A(\varepsilon)$ 有左 Perron 向量 $y(\varepsilon)>0$,

$$y(\varepsilon)^{\mathrm{T}}A(\varepsilon)=\rho(A(\varepsilon))y(\varepsilon)^{\mathrm{T}}.$$

由于 $Ax-\alpha x\geq 0$,有

$$A(\varepsilon)x-\alpha x\geq Ax-\alpha x\geq 0,$$

因此

$$y(\varepsilon)^{\mathrm{T}}\left(A(\varepsilon)x-\alpha x\right)=\left(\rho\left(A(\varepsilon)\right)-\alpha\right)y(\varepsilon)^{\mathrm{T}}x\geq 0.$$

注意到 $y(\varepsilon)^{\mathrm{T}}x>0$,得

$$\rho(A(\varepsilon))-\alpha\geq 0,\quad \forall\varepsilon>0.$$

由此及当 $\varepsilon\to 0$ 时 $\rho(A(\varepsilon))\to\rho(A)$,推出 $\rho(A)\geq\alpha$.

"$Ax>\alpha x$ 时 $\rho(A)>\alpha$",即 **4.2.2**,已作论证. □

这一定理是 **4.1.11** 的延伸.

4.4.5 推论 设 $A\in\mathbb{R}^{n\times n}$, $A\geq 0$,则

$$\rho(A)=\max_{x\geq 0,x\neq 0}\min_{1\leq i\leq n,x_i\neq 0}\frac{1}{x_i}\sum_{j=1}^{n}a_{ij}x_j,$$

其中记号 $x=(x_1,\cdots,x_n)^{\mathrm{T}}$.

证 设 $x\geq 0$, $x\neq 0$,取

$$\alpha\equiv\min_{1\leq i\leq n,x_i\neq 0}\frac{1}{x_i}\sum_{j=1}^{n}a_{ij}x_j,$$

则 $Ax\geq\alpha x$.依 **4.4.4**,$\alpha\leq\rho(A)$.然而,如果选如此 x 为 A 的特征向量,**4.4.2** 保证了这种特征向量的存在性,那么上界即

$$\alpha=\rho(A)$$

是可以达到的. □

4.5 随机矩阵和双随机矩阵

4.5.1 定义 设 $S=[s_{ij}]\in\mathbb{R}^{n\times n}$, $S\geq 0$ 且满足

$$\sum_{i=1}^{n} s_{ij} = 1, \quad j = 1,\cdots,n, \tag{5.1}$$

则称 S 为**随机矩阵**(stochastic matrix).如果还满足

$$\sum_{j=1}^{n} s_{ij} = 1, \quad i = 1,\cdots,n, \tag{5.2}$$

则称 S 为**双随机矩阵**(doubly stochastic matrix/bistochastic matrix).

一个具体的解释是研究 n 个种类的采集,其每一种类变更成另一种类的可能性.数 s_{ij} 称为转移概率,表示第 i 种类变更成第 j 种类的总体分数.

在如此解释中,$S \geq 0$ 是自然的;条件(5.1)和(5.2)则指明总体保持一定.可以经历变更的种类的类型是原子核,转变共享公共生态环境,以及其它许多情形.

在有些重要应用中则另有解释.

显然,每个置换矩阵是随机矩阵和双随机矩阵.

4.5.2 定理 设 $S \in \mathbb{R}^{n\times n}$ 是正随机矩阵,则

(1) S 的谱半径 $\rho(S) = 1$.

(2) 对于任一非零非负向量 z,成立

$$\lim_{k\to\infty} S^k z = cx, \tag{5.3}$$

其中 x 是相应 $\rho(S)$ 的特征向量,c 是某正常数.

证 注意到 S^{T} 和 S 有相同的特征值,可以转而考虑 S^{T} 的谱半径.根据条件(5.1),显然

$$\xi = (1,\cdots,1)^{\mathrm{T}}$$

是 S^{T} 的特征向量,而且 1 是 S^{T} 的特征值,ξ 相应于特征值 1.

另一方面,由于 S 是正矩阵,S^{T} 也是正矩阵,又 $\xi > 0$,依 **4.2.4**,ξ 是 S^{T} 的优势特征向量,即相应优势特征值 $\rho(S^{\mathrm{T}})$ 的特征向量.因此

$$\rho(S)=\rho(S^{\mathrm{T}})=1.$$

(1)得证.

现在证明(2).先假定 S 存在 n 个线性无关的真特征向量

$$\boldsymbol{x}^{(1)},\cdots,\boldsymbol{x}^{(n)}.$$

将 z 表示成 $\boldsymbol{x}^{(1)},\cdots,\boldsymbol{x}^{(n)}$ 的线性组合:

$$z=\sum_{i=1}^{n}c_i x^{(i)}. \tag{5.4}$$

于是

$$S^k z=\sum_{i=1}^{n}c_i\lambda_i^k x^{(i)}, \tag{5.5}$$

其中 λ_i 是对应于 $\boldsymbol{x}^{(i)}$ 的特征值.不妨假定

$$\lambda_1=\rho(S)=1,$$

$x^{(1)}>0$ 是 S 的属于 $\rho(S)$ 的优势特征向量.依 **4.2.3**,

$$|\lambda_i|<1,\quad i=2,\cdots,n.$$

于是从(5.5)推出(5.3),其中 $c=c_1, x=\boldsymbol{x}^{(1)}$.因为

$$\xi=S^{\mathrm{T}}\xi=\left(S^{\mathrm{T}}\right)^k\xi,$$

由此及(5.3)得

$$c(x,\xi)=\lim_{k\to\infty}\left(S^k z,\xi\right)=\lim_{k\to\infty}\left(z,\left(S^{\mathrm{T}}\right)^k\xi\right)=(z,\xi). \tag{5.6}$$

由于 $\xi>0, x>0$,并根据假设,$z\geq 0, z\neq 0$,从(5.6)推出 $c>0$.这样,对于存在 n 个真特征向量的情形证明了(2).

关于存在广义特征向量的一般情形,可按类似的方式推证,留作练习. □

应用上,考虑由转移概率控制变更的系统.

用 x_j 表示第 j 种类的总体大小,$j=1,\cdots,n$.假设在单位时间(1 年或 1 天或 1×10^{-9} 秒等)内采集的每个个体按概率 s_{ij} 变更(或增殖)为其它种类的成员,而且假定总体大小非常大,以致波动是无关

紧要的.那么,第i种类新总体大小是

$$y_i=\sum_{j=1}^{n}s_{ij}x_j. \tag{5.7}$$

将老总体分量和新总体分量分别组合成单个列向量x和y,由(5.7)有

$$y=Sx. \tag{5.8}$$

经k个单位时间后,总体向量为$S^k x$.在如此应用中,上述定理的意义是当$k\to\infty$时,该总体趋向于与总体从何处起始无关的定常分布.

4.5.3 定理 设$S\in\mathbb{R}^{n\times n}$是双随机矩阵,则

(1) $\rho(S)=1$,且$\xi=(1,\cdots,1)^{\mathrm{T}}$是相应的特征向量.

(2) $\|S\|_2\geq 1$.

(3) S为矩阵的充分必要条件是S有n个非零元素.

证明非常容易,留作练习.

4.5.4 Birkhoff 定理 $n\times n$双随机矩阵全体组成的集合是$n\times n$置换矩阵全体的凸包,就是说,任一$n\times n$双随机矩阵$S=[s_{ij}]$可表示成$n\times n$置换矩阵$P_i\ (i=1,\cdots,n!)$的凸组合:

$$S=\sum_{i=1}^{n!}\alpha_i P_i,\quad \sum_{i=1}^{n!}\alpha_i=1,\quad \alpha_i\geq 0,\quad i=1,\cdots,n! \tag{5.9}$$

证 对S的非零元素个数$\nu(S)\geq n$用数学归纳法.

当$\nu(S)=n$时,S就是置换矩阵,定理成立.

现设$\nu(S)>n$.假定定理对非零元素个数少于$\nu(S)$的任何$n\times n$双随机矩阵成立,要证明定理对S也成立.

由于$\nu(S)>n$,必存在行指标i_1和列指标j_1,

$$0<s_{i_1 j_1}<1.$$

由此,又必存在行指标$i_2\neq i_1$,而后存在列指标$j_2\neq j_1$,

$$0 < s_{i_2 j_1} < 1, \quad 0 < s_{i_2 j_2} < 1.$$

如此继续,因只有 n 行,必出现行指标 i_{m+1} 与前面某行指标 i_l 相重,即 $i_{m+1} = i_l$,元素选取过程至此为止.

不妨设 $l = 1$,否则改为从第 i_l 行出发,将其改记 i_1,再对其后者 $i_{l+1}, \cdots, i_{m+1}$ 加以重新编号即可.这样,选择了 $2m$ 个元素满足如下条件:

$$0 < s_{i_k j_k}, s_{i_{k+1} j_k} < 1, \quad k = 1, \cdots, m. \tag{5.10}$$

将它们的所在位置分成两组:

$$\mathscr{L}_1 \equiv \{(i_k, j_k): k = 1, \cdots, m\}$$

和

$$\mathscr{L}_2 \equiv \{(i_{k+1}, j_k): k = 1, \cdots, m\},$$

注意,依以上所设,$i_{m+1} = i_1$.如此选取的 $2m$ 个元素,从涉及的第 $i_1, \cdots,$ 第 i_m 这 m 个行看,每行占 2 个;从涉及的第 $j_1, \cdots,$ 第 j_m 这 m 个列看,每列也占 2 个.于是,只要取

$$\hat{s} = \min_{(i,j) \in \mathscr{L}_1} s_{ij}, \quad \tilde{s} = \min_{(i,j) \in \mathscr{L}_2} s_{ij},$$

然后令

$$\hat{s}_{ij} = \begin{cases} s_{ij} - \hat{s}, & (i,j) \in \mathscr{L}_1, \\ s_{ij} + \hat{s}, & (i,j) \in \mathscr{L}_2, \\ s_{ij}, & (i,j) \notin \mathscr{L}_1 \cup \mathscr{L}_2, \end{cases}$$

和

$$\tilde{s}_{ij} = \begin{cases} s_{ij} + \tilde{s}, & (i,j) \in \mathscr{L}_1, \\ s_{ij} - \tilde{s}, & (i,j) \in \mathscr{L}_2, \\ s_{ij}, & (i,j) \notin \mathscr{L}_1 \cup \mathscr{L}_2, \end{cases}$$

则因 S 是双随机矩阵,得知

$$\hat{S} \equiv [\hat{s}_{ij}], \quad \tilde{S} \equiv [\tilde{s}_{ij}],$$

也都是双随机矩阵,而且

$$\nu(\hat{S})<\nu(S), \quad \nu(\tilde{S})<\nu(S).$$

由归纳法假设,有

$$\hat{S}=\sum_{i=1}^{n!}\hat{\alpha}_i P_i, \quad \sum_{i=1}^{n!}\hat{\alpha}_i=1, \quad \hat{\alpha}_i\geq 0, \quad i=1,\cdots,n!$$

和

$$\tilde{S}=\sum_{i=1}^{n!}\tilde{\alpha}_i P_i, \quad \sum_{i=1}^{n!}\tilde{\alpha}_i=1, \quad \tilde{\alpha}_i\geq 0, \quad i=1,\cdots,n!$$

由此推出

$$S=\frac{\tilde{s}}{\hat{s}+\tilde{s}}\hat{S}+\frac{\hat{s}}{\hat{s}+\tilde{s}}\tilde{S}=\sum_{i=1}^{n!}\alpha_i P_i,$$

其中

$$\alpha_i=\frac{\tilde{s}}{\hat{s}+\tilde{s}}\hat{\alpha}_i+\frac{\hat{s}}{\hat{s}+\tilde{s}}\tilde{\alpha}_i, \ i=1,\cdots,n!$$

而且

$$\sum_{i=1}^{n!}\alpha_i=\frac{\tilde{s}}{\hat{s}+\tilde{s}}\sum_{i=1}^{n!}\hat{\alpha}_i+\frac{\hat{s}}{\hat{s}+\tilde{s}}\sum_{i=1}^{n!}\tilde{\alpha}_i .$$

至此,归纳法步骤已完成. □

4.5.5 定理(Dénes König & Garrett Brikhoff) $n\times n$ 置换矩阵是 $n\times n$ 双随机矩阵全体所成之集合的极值点.反之亦然.

证 由 **4.5.4** 可知,$n\times n$ 双随机矩阵全体所成之集合 K 是$\mathbb{R}^{n\times n}$ 中的有界闭凸集.

先证所有置换矩阵是 K 的极值点(见 **1.9.14**).设 $P=[p_{ij}]$ 是任一 $n\times n$ 置换矩阵,假定

$$P=\frac{A+B}{2}, \quad A=[a_{ij}],B=[b_{ij}]\in K,$$

由此推出:若 $p_{ij}=1$,则相应地有

$$a_{ij}=b_{ij}=1.$$

若 $p_{ij}=0$,则相应地有

$$a_{ij}=b_{ij}=0,$$

因此

$$\boldsymbol{P}=\boldsymbol{A}=\boldsymbol{B},$$

故 P 是 K 的极值点.

再证非置换矩阵的双随机矩阵必不是 K 的极值点.设双随机矩阵 $S=[s_{ij}]$ 不是置换矩阵,亦即存在元素 $s_{i_0j_0}$,

$$0<s_{i_0j_0}<1,$$

则根据双随机矩阵的定义,由 $s_{i_0j_0}$ 出发可以构造 S 元素的如下序列:

选取列指标 j_1,使得 $0<s_{i_0j_1}<1$;

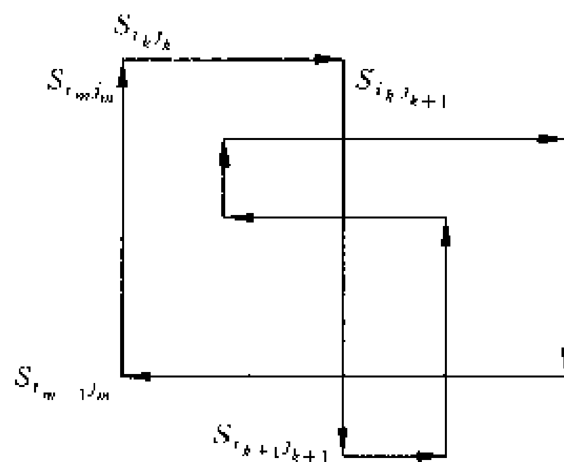
选取行指标 i_1,使得 $0<s_{i_1j_1}<1$;

……

一直继续至同一位置第二次被选中.如此,可以得到一个闭链:

$$s_{i_kj_k}\to s_{i_kj_{k+1}}\to\cdots\to s_{i_mj_m}=s_{i_kj_k}.$$

示意图如下:



现在定义矩阵 N 如下:

(1) N 的不在链的位置上的元素全为 0;

(2) N 的相应于链的位置上的元素逐次交替取 $+1$ 和 -1;

从而有

(3) N 的每行之和与每列之和均为 0.

然后,再定义

$$A \equiv S + \varepsilon N, \quad B \equiv S - \varepsilon N.$$

从(3)推出 A 与 B 两者的每行之和与每列之和均为 1.从(1)及(2)得知 S 的相应 N 的非零元素位置的元素都是正的.因此,可以选取小的 ε,使得 A 与 B 都是非负的,成为双随机矩阵.由于

$$S = \frac{A+B}{2}, \quad A \neq B,$$

这表明 S 不是 K 的极值点. □

4.5.6 定理 设 $A = [a_{ij}] \in \mathbb{C}^{n\times n}$ 是正规矩阵,它的谱分解为

$$A = U\Lambda U^*,$$

其中 $U = [u_{ij}] \in \mathbb{C}^{n\times n}$ 是酉矩阵,$\Lambda = \operatorname{diag}(\lambda_1, \cdots, \lambda_n)$,则 A 的主对角元素构成的向量

$$d(A) \equiv (a_{11}, \cdots, a_{nn})^{\mathrm{T}} \in \mathbb{C}^n,$$

和 A 的特征值构成的向量

$$e(A) \equiv (\lambda_1, \cdots, \lambda_n)^{\mathrm{T}} \in \mathbb{C}^n,$$

存在如下线性变换关系:

$$d(A) = Se(A), \tag{5.11}$$

其中 S 是矩阵 U 和 $\overline{U}$ 的 Hadamard 积(见 **7.2.1**):

$$S \equiv \left[\, |u_{ij}|^2 \right] = \left[U \circ \overline{U}\right] \in \mathbb{R}^{n\times n}, \tag{5.12}$$

而且是双随机矩阵.

证 从 $A = U\Lambda U^*$ 直接计算有

$$\begin{aligned} a_{ii} &= \lambda_1 u_{i1}\overline{u_{i1}} + \cdots + \lambda_n u_{in}\overline{u_{in}} \\ &= |u_{i1}|^2 \lambda_1 + \cdots + |u_{in}|^2 \lambda_n, \quad i = 1, \cdots, n, \end{aligned}$$

由此即得(5.11).因为 $UU^* = I$,显然 S 是双随机矩阵. □

4.5.7 定义 设$x=(x_1,\cdots,x_n)^{\mathrm{T}},y=(y_1,\cdots,y_n)^{\mathrm{T}}\in\mathbb{R}^n$，用

$$x_{[1]}\geq\cdots\geq x_{[n]}\quad 和\quad y_{[1]}\geq\cdots\geq y_{[n]} \tag{5.13}$$

分别表示x和y的元素按代数递减顺序重新编号;用

$$x_{(1)}\leq\cdots\leq x_{(n)}\quad 和\quad y_{(1)}\leq\cdots\leq y_{(n)} \tag{5.14}$$

分别表示x和y的元素按代数递增顺序重新编号.如果满足

$$\sum_{i=1}^{k}x_{[i]}\leq\sum_{i=1}^{k}y_{[i]},\quad k=1,\cdots,n, \tag{5.15}$$

则称y**弱优化**于x.如果满足(5.15)且当$k=n$时成立等号,亦即

$$\begin{gathered}\sum_{i=1}^{k}x_{[i]}\leq\sum_{i=1}^{k}y_{[i]},\quad k=1,\cdots,n-1;\\ \sum_{i=1}^{n}x_{[i]}=\sum_{i=1}^{n}y_{[i]}\end{gathered} \tag{5.16}$$

此时等价于满足

$$\begin{gathered}\sum_{i=1}^{k}x_{(i)}\geq\sum_{i=1}^{k}y_{(i)},\quad k=1,\cdots,n-1;\\ \sum_{i=1}^{n}x_{(i)}=\sum_{i=1}^{n}y_{(i)},\end{gathered} \tag{5.17}$$

则称y**(强)优化**于x.

4.5.8 定理 设

$$x=(x_1,\cdots,x_n)^{\mathrm{T}},y=(y_1,\cdots,y_n)^{\mathrm{T}}\in\mathbb{R}^n,x\geq0,y\geq0,$$

则y(强)优化于x的充分必要条件是存在双随机矩阵$S\in\mathbb{R}^{n\times n}$,使得$x=Sy$.

证 若$S=[s_{ij}]\in\mathbb{R}^{n\times n}$是双随机矩阵,$x=Sy$，则有

$$\sum_{i=1}^{k}x_{[i]}=\sum_{i=1}^{k}(Sy)_{[i]}=\sum_{i=1}^{k}\left(\sum_{j=1}^{n}s_{[i]j}y_j\right)=\sum_{i=1}^{k}\left(\sum_{j=1}^{n}s_{[i][j]}y_{[j]}\right)$$

$$\le \sum_{i=1}^{k}\left[\sum_{j=1}^{k-1}s_{[i][j]}y_{[j]}+\left(\sum_{j=k}^{n}s_{[i][j]}\right)y_{[k]}\right]$$

$$=\sum_{i=1}^{k}\left[\sum_{j=1}^{k-1}s_{[i][j]}y_{[j]}+\left(1-\sum_{j=1}^{k-1}s_{[i][j]}\right)y_{[k]}\right]$$

$$=\sum_{i=1}^{k}\left[\sum_{j=1}^{k-1}s_{[i][j]}\left(y_{[j]}-y_{[k]}\right)+y_{[k]}\right]$$

$$=\sum_{j=1}^{k-1}\left(\sum_{i=1}^{k}s_{[i][j]}\right)\left(y_{[j]}-y_{[k]}\right)+ky_{[k]}$$

$$\le \sum_{j=1}^{k-1}\left(y_{[j]}-y_{[k]}\right)+ky_{[k]}=\sum_{j=1}^{k}y_{[j]},$$

$$k=1,\cdots,n-1.$$

而且

$$\sum_{i=1}^{n}x_{[i]}=\sum_{i=1}^{n}(Sy)_{[i]}=\sum_{i=1}^{n}\left(\sum_{j=1}^{n}s_{ij}y_{j}\right)_{[i]}=\sum_{i=1}^{n}\sum_{j=1}^{n}s_{ij}y_{j}$$

$$=\sum_{j=1}^{n}\left(\sum_{i=1}^{n}s_{ij}\right)y_{j}=\sum_{j=1}^{n}y_{j}=\sum_{j=1}^{n}y_{[j]},$$

因此 y 优化于 x.充分性得证.

必要性的证明留作练习. □

4.6 M-矩阵

4.6.1 定义 用 $Z_n\subset\mathbb{R}^{n\times n}$ 表示非对角元素非正的 $n\times n$ 实矩阵的集合:

$$Z_n \equiv \left\{A=[a_{ij}]\in\mathbb{R}^{n\times n}: a_{ij}\le 0, i\ne j,\quad i,j=1,\cdots,n\right\}. \tag{6.1}$$

而且,若 $A\in Z_n$,则称 A 为 **Z-矩阵**.

显然,$A=[a_{ij}]\in Z_n$ 的充分必要条件是 $A\in\mathbb{R}^{n\times n}$ 而且对于某 $P\in\mathbb{R}^{n\times n}, P\ge 0$,以及某 $\alpha\in\mathbb{R}$,使得

$$A=\alpha I-P, \tag{6.2}$$

α 和 P 的选取有很大的任意性,例如取 $\alpha\ge\max\limits_{1\le i\le n} a_{ii}$,还有总可以取 $\alpha>0$ 等.

Z_n 中矩阵的表示形式(6.2)很有好处,因为它启示着与非负矩阵的 Perron-Frobenius 理论相联系.

4.6.2 引理 设 $A=[a_{ij}]\in Z_n$,并设有表示式

$$A=\alpha I-P,\quad \alpha\in\mathbb{R},\ P\ge 0,$$

则

$$\alpha-\rho(P)$$

是 A 的特征值,而且

$$\lambda(A)\subset D\equiv\left\{z\in\mathbb{C}:\left|z-\alpha\right|\le\rho(P)\right\}, \tag{6.3}$$

并因此 A 的每个特征值 λ 满足

$$\operatorname{Re}\lambda\ge\alpha-\rho(P). \tag{6.4}$$

特别,A 正稳定的充分必要条件是

$$\alpha>\rho(P). \tag{6.5}$$

若 A 是正稳定的,则可写成

$$A=\gamma I-P,\quad \gamma=\max_{1\le i\le n}a_{ii},\quad P=\gamma I-A\ge 0, \tag{6.6}$$

且必有 $\gamma>\rho(P)$.

证 设 $A=\alpha I-P, \alpha\in\mathbb{R}, P\ge 0$.注意到 A 的每个特征值形如

$$\alpha - e(P),$$

其中 $e(P)$ 是 P 的特征值.因此,A 的每个特征值在复平面中心为 $z=\alpha$ 半径为 $\rho(P)$ 的圆盘之中,即(6.3)成立.

因 P 非负,$\rho(P)$ 是其特征值,故 $\alpha - \rho(P)$ 是 A 的实特征值.

若 A 是正稳定的,则 $\alpha - \rho(P) > 0$,即(6.6)成立;反之,若 $\alpha > \rho(P)$,则圆盘 D 在右半平面,因此 A 是正稳定的. A 正稳定时,表示形式(6.6)和 $\gamma > \rho(P)$ 是显然的. □

4.6.3 定义 如果 $A=[a_{ij}] \in Z_n$,而且满足

$$a_{ii} > 0, \quad i = 1, \cdots, n,$$

则称 A 为 **L-矩阵**或 **Metzler 矩阵**.

4.6.4 定义 为了方便应用,引进非奇异 M-矩阵的三种定义如下:

M-矩阵定义之一: 如果 $A=[a_{ij}] \in Z_n$ 是非奇异的,而且 $A^{-1} \geq 0$,则称 A 为**非奇异 M-矩阵**(以 Minkowski 命名).

M-矩阵定义之二: 如果 $A \in Z_n$ 可表示为

$$A = \alpha I - P,$$

其中 $P \geq 0$ 且 $\alpha > \rho(P)$,则称 A 为**非奇异 M-矩阵**.

M-矩阵定义之三: 如果 $A \in Z_n$ 是正稳定的,则称 A 为**非奇异 M-矩阵**.

例 矩阵

$$A = \begin{bmatrix} 2 & -1 & 0 \\ -1 & 2 & -1 \\ 0 & -1 & 2 \end{bmatrix}, \tag{6.7}$$

显然 $A \in Z_3$.因为 $\det A \neq 0$,A 是非奇异的,而且

$$A^{-1}=\frac{1}{4}\begin{bmatrix}3 & 2 & 1\\ 2 & 4 & 2\\ 1 & 2 & 3\end{bmatrix}>0,$$

A 符合 M-矩阵定义之一.

A 可表示为 $A=\alpha I-P$,

$$\alpha=2,\quad P=\begin{bmatrix}0 & 1 & 0\\ 1 & 0 & 1\\ 0 & 1 & 0\end{bmatrix},$$

P 的特征值是 $\sqrt{2}$,0 ,$-\sqrt{2}$, $\rho(P)=\sqrt{2}<2=\alpha$. A 符合 M-矩阵定义之二.

A 的特征值是 $\lambda_1=2+\sqrt{2}$, $\lambda_2=2$, $\lambda_3=2-\sqrt{2}$,它们都是正的,故 A 是正稳定矩阵. A 符合 M-矩阵定义之三.

一般, $n\times n$ 实三对角矩阵

$$A=\begin{bmatrix}a_1 & b_1 & & & 0\\ c_1 & a_2 & b_2 & & \\ & \ddots & \ddots & \ddots & \\ & & \ddots & \ddots & b_{n-1}\\ 0 & & & c_{n-1} & a_n\end{bmatrix}\tag{6.8}$$

如果 A 元素满足

$$a_i>0,\quad i=1,\cdots,n;\quad b_i<0,c_i<0,\quad i=1,\cdots,n-1;$$

而且其各行之和满足

$$a_1+b_1>0,\quad a_n+c_{n-1}>0;$$

$$a_i+b_i+c_{i-1}\geq 0,\quad i=2,\cdots,n-1,$$

则 A 是 M-矩阵.

M-矩阵出现在许多应用领域,是一类重要的矩阵.以上三种定

义分别为一些教科书和著作所采用.定义之一比较常见,定义之二显示与非负矩阵的联系,定义之三在一些场合很好用,各有所长.下面的定理表明它们是等价的,而且M-矩阵的定义还存在很多其它的选择.

4.6.5 定理 对于 $A=[a_{ij}]\in Z_n$,如下各条件是等价的:

(1) A 是非奇异的,而且 $A^{-1}\ge 0$.

(2) 若 $Ax\ge 0$,则 $x\ge 0$.

(3) $A=\alpha I-P$, $P\ge 0$, $\alpha>\rho(P)$.

(4) A 是L-矩阵,而且 $A^{-1}\ge 0$.

(5) A 的每个实特征值是正的.

(6) 对于任意的 $t\ge 0$, $A+tI$ 是非奇异的.

(7) 对于每个非负对角矩阵 D, $A+D$ 是非奇异的.

(8) A 的所有主子式是正的.

(9) A 的所有 k 阶主子式——共 $\binom{n}{k}$ 个——之和 $E_k(A)$ 是正的, $k=1,\cdots,n$.

(10) A 的前主子式是正的.

(11) $A=LU$,其中 L 是下三角矩阵, U 是上三角矩阵,而且两者的所有对角元素是正的.

(12) 对于每个非零向量 $x=(x_1,\cdots,x_n)^{\mathrm{T}}\in\mathbb{R}^n$,存在指标 i, $1\le i\le n$,使得

$$x_i(a_{i1}x_1+a_{i2}x_2+\cdots+a_{in}x_n)>0$$

(13) 对于每个非零向量 $x\in\mathbb{R}^n$,存在**正对角矩阵**——对角元素全为正实数的对角矩阵——D,使得 $x^{\mathrm{T}}ADx>0$.

(14) 存在正向量 $x\in\mathbb{R}^n$,使得 $Ax>0$.

(15) A 是L-矩阵,而且是拟严格对角优势矩阵.

一般地说, $A\in\mathbb{C}^{n\times n}$ 称为**拟严格对角优势矩阵**(quasi strictly diagonally dominant matrix),如果存在正对角矩阵 D,使得 AD 是严格对角优势的.

(16) A 是 L-矩阵,而且存在正对角矩阵 D,使得 $D^{-1}AD$ 是行严格对角优势的.

(17) A 是 L-矩阵,而且存在正对角矩阵 D 和 E,使得 DAE 同时为行严格对角优势和列严格对角优势.

(18) 存在正对角矩阵 D,使得 $DA+A^{\mathrm{T}}D$ 是正定矩阵;也就是说,存在正对角 Lyapunov 解.

(19) A 是正稳定矩阵.

证 (1)⇒(2): 设 $Ax \equiv y \geq 0$,则由 $A^{-1} \geq 0$ 推出

$$x = A^{-1}y \geq 0.$$

(2)⇒(3): 将 A 写成

$$A = \alpha I - P,\ \ P \geq 0,\ \ \alpha > 0.$$

设 v 是 P 的 Perron 右特征向量:

$$v \geq 0,\ \ v \neq 0,\ \ Pv = \rho(P)v.$$

如果 $\alpha \leq \rho(P)$,那么

$$A(-v) = (\rho(A) - \alpha)v \geq 0.$$

这与条件(2)矛盾.因此,$\alpha > \rho(P)$.

(3)⇒(1): 因为除以正数 α 对条件(3)不会有重要的变化,不妨假设 $\alpha = 1$.如此,因 $\rho(P) < 1$,幂级数 $I + P + P^2 + \cdots$ 收敛;而且因 $P \geq 0$,

$$I + P + P^2 + \cdots = (I - P)^{-1} = A^{-1} \geq 0.$$

(3)⇒(4): 以上(3)⇒(1)的证明中得知(3)蕴涵 A 非奇异且 $A^{-1} \geq 0$.现设 $A^{-1} \equiv [b_{ij}]$,由 $A^{-1}A = I$ 有

$$\sum_{k=1}^{n} b_{ik}a_{ki} = 1,\quad i = 1,\cdots,n.$$

由此并注意到 $a_{ij} \leq 0, i \neq j$,有

$$b_{ii}a_{ii} = 1 - \sum_{k=1,k\neq i}^{n} b_{ik}a_{ki} \geq 1,\quad i = 1,\cdots,n.$$

于是,从 $b_{ii} \geq 0$ 推出 $a_{ii} > 0, i = 1,\cdots,n$.

(4)$\Rightarrow$(3): 已知

$$a_{ij} \le 0,\ i \ne j;\quad a_{ii} > 0,\ i = 1,\cdots,n.$$

而且有 $A^{-1} \ge 0$.令

$$\alpha \ge \max\{a_{11},\cdots,a_{nn}\},$$

则有

$$P \equiv \alpha I - A \ge 0.$$

于是 $\rho(P)$是 P 的特征值,而且存在对应的特征向量 $x \ge 0, x \ne 0$,

$$Px = \alpha x - Ax = \rho(P)x.$$

因此

$$Ax = (\alpha - \rho(P))x,$$

从而有

$$(\alpha - \rho(P))A^{-1}x = x \ge 0,$$

由此推出 $\alpha > \rho(P)$.

(3)$\Leftrightarrow$(19): 见 **4.6.2**.

(19)$\Rightarrow$(5): 正稳定矩阵定义 **3.14.3** 蕴涵每个实特征值是正的.

(5)$\Rightarrow$(3): $A \in Z_n$ 可以写成 $A = \alpha I - P, P \ge 0$, 而且有 $\alpha - \rho(P) \in \lambda(A)$.(5)蕴涵 $\alpha - \rho(P) > 0$,即 $\alpha > \rho(P)$.

(3)$\Rightarrow$(6): 由(3)有

$$A + tI = (\alpha + t)I - P,\quad \alpha + t > \rho(P),\quad \forall t > 0,$$

这表明用 $A + tI$ 代替 A 时(3)成立.

因(3)蕴涵(19),故 $A + tI \in \mathbb{R}^{n\times n}$ 是正稳定的.依 **3.14.5**,有

$$\det(A + tI) > 0,$$

因此当 $t > 0$ 时 $A + tI$ 是非奇异的.

(3)$\Rightarrow$(8): 容易看出,条件(3)可以遗传给 A 的主子矩阵,而且因(3)蕴涵(19),推出 A 的每个主子矩阵是正稳定的.注意到 $A \in \mathbb{R}^{n\times n}$,依 **3.14.5**,这些主子矩阵的行列式大于零.

(9)$\Rightarrow$(6): $\det(A + \alpha I)$可展开成首项系数为 1 的 t 的 n 次多项式,其 t^{n-k} 的系数是 $E_k(A)$.(9)保证

$$E_k(A)>0,\quad k=1,\cdots,n,$$

于是

$$\det(A+tI)>0,\quad \forall t\geq 0.$$

(10)⇒(11): 从 **5.8.4** 的 LU 分解推证中即可得出.

(3)⇒(14): 若 $P\geq 0$ 不可约,设 $x>0$ 是 P 的 Perron 右特征向量;若 P 可约,在 P 中为零的那些位置加以扰动,替换为充分小的值 $\varepsilon>0$,并设 $x>0$ 是所得矩阵的 Perron 右特征向量.因此在 $Ax=\alpha x-Px$ 中,Px 或者等于 $\rho(P)x$,或者接近于 $\rho(P)x$ 到任意程度.无论哪种情形,Ax 充分接近于 $(\alpha-\rho(P))x>0$,故必有 $Ax>0$.

(14)⇒(15): 已知存在 $x=(x_1,\cdots,x_n)^{\mathrm{T}}>0$,使得 $Ax>0$;取 $D=\mathrm{diag}(x_1,\cdots,x_n)$.注意到 AD 的各行之和正好是

$$ADe=Ax>0$$

的元素,其中 $e\equiv(1,\cdots,1)^{\mathrm{T}}\in\mathbf{R}^n$.因为 $A\in Z_n$ 的所有非对角元素是非正的,所以 AD 的对角元素并因而 A 的对角元素必须是正的,而且 AD 必须是行严格对角优势的.

(17)⇒(18): 若(17)成立,则

$$DAE+(DAE)^{\mathrm{T}}$$

有正对角元素,是严格对角优势的,并因此是正定的.用 E^{-1} 作合同变换,表明

$$E^{-1}\left(DAE+(DAE)^{\mathrm{T}}\right)E^{-1}=\left(E^{-1}D\right)A+A^{\mathrm{T}}\left(E^{-1}D\right)$$

也是正定的.这就是说,$E^{-1}D$ 是正对角 Lyapunov 解.

(18)⇒(8): 注意到条件(18)可以遗传给 A 的主子矩阵,并且依 **3.14.6**,(18)蕴涵正稳定性.因此,A 的每个主子矩阵是实正稳定矩阵,所以行列式大于零,亦即(8)成立.

(13)⇒(19): 这是 **3.14.17** 的直接推论.

以上已完成条件(1),(2),(3),(4),(5),(19)之间等价性的证明;其余的只给出部分证明,而对(7),(12),(16)则未涉及;未给出的证明大多是

比较容易的,留作练习. □

应当注意,定理中大多数条件在 Z_n 之外是不等价的. Z_n 所赋予的结构是值得重视的,其最主要的原由是蕴涵了与非负矩阵的关系.

M-矩阵存在大量的充分必要条件,以上定理中所列的只是筛选出来的一部分,它们对认识和应用来源于各种渠道的M-矩阵是有用的.

M-矩阵和正定矩阵之间存在非常多的类似的和平行的结论.例如,以上定理中的等价条件(8),(9),(10).下面两个定理进一步表明了这一点.

练习 根据以上定理条件(11), M-矩阵可以 LU 分解.证明 L 和 U 都是M-矩阵.反之,若 L 和 U 都是M-矩阵,试问其积 LU 是M-矩阵吗?

4.6.6 定理 设 $A=[a_{ij}],B=[b_{ij}]\in Z_n$, A 是 M-矩阵,且 $B\geq A$,则

(1) B 是M-矩阵.

(2) $A^{-1}\geq B^{-1}\geq 0$.

(3) $\det B\geq \det A>0$.

证 (1)从 **4.6.5** 的(14)直接推出.因为若 $x>0$ 使得 $Ax>0$,则 $Bx\geq Ax>0$.

现在, A 和 B 都是 M-矩阵,有 $A^{-1}\geq 0$ 和 $B^{-1}\geq 0$.由于

$$A^{-1}-B^{-1}=A^{-1}(B-A)B^{-1},$$

而等式右端是三个非负矩阵之积,因此

$$A^{-1}-B^{-1}\geq 0.$$

(2)得证.

为了证明(3),对维数 n 应用归纳法.

要推证的不等式对于 $n=1$ 显然成立.

现设(3)对维数$k=1,\cdots,n-1$全成立.将A和B分块如下:

$$A=\begin{bmatrix} A_{11} & A_{12} \\ A_{21} & a_{nn} \end{bmatrix}, \quad B=\begin{bmatrix} B_{11} & B_{12} \\ B_{21} & b_{nn} \end{bmatrix},$$

显然,$A_{11},B_{11}\in Z_{n-1}$,$B_{11}\ge A_{11}$;而且,依 **4.6.5**,主子矩阵A_{11}和B_{11}是M-矩阵.由归纳法假设有

$$\det B_{11}\ge \det A_{11}>0.$$

再注意到还有

$$A^{-1}\ge B^{-1}\ge 0,\ \det A>0,\ \det B>0,$$

因此,非负矩阵$A^{-1}-B^{-1}$的(n,n)位置的元素

$$\left(A^{-1}-B^{-1}\right)_{nn}=\frac{\det A_{11}}{\det A}-\frac{\det B_{11}}{\det B}\ge 0,$$

于是

$$\det B\ge\left(\frac{\det B_{11}}{\det A_{11}}\right)\det A\ge \det A>0. \qquad \square$$

4.6.7 定理 设$A=[a_{ij}]\in Z_n$是M-矩阵.则成立

(1) **Hadamard 不等式**

$$\det A\le a_{11}\cdots a_{nn}.$$

(2) **Fischer 不等式**

$$\det A\le \det A(\alpha)\det A(\alpha'),\ \forall\alpha\subseteq\{1,\cdots,n\},$$

其中α和$\alpha'\equiv\{1,\cdots,n\}-\alpha$是指标集,$A(\alpha)$表示$A$取$\alpha$所含指标的行和列确定的主子矩阵.

(3) **Szasz 不等式**

$$P_{k+1}^{\binom{n-1}{k}^{-1}}\le P_k^{\binom{n-1}{k-1}^{-1}},\quad k=1,\cdots,n-1,$$

其中P_k表示A的所有k阶主子式之积.

证 显然,(1)是(2)的推论.

现在证明(2).不失一般性,假定

$$\alpha = \{1, \cdots, k\}, \quad 1 \le k \le n.$$

并且将 A 分块如下:

$$A = \begin{bmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{bmatrix},$$

其中 $A_{11} \in Z_k$, $A_{22} \in Z_{n-k}$.定义

$$B = \begin{bmatrix} A_{11} & 0 \\ 0 & A_{22} \end{bmatrix},$$

则 $B \in Z_n$, $B \ge A$.依 **4.6.6** 的(3),

$$\det A \le \det B = \det A_{11} \det A_{22},$$

Fischer 不等式得证.

Szasz 不等式的证明可参见[HJ1985]的(7.8.2). □

4.6.8 定理 设 $A \in Z_n$ 是对称的,则 A 为非奇异 M-矩阵的充分必要条件是 A 为正定矩阵.

证 依 **4.6.5**, $A \in Z_n$ 为非奇异 M-矩阵等价于 A 的前主子式是正的;而在 A 是实对称的条件下,依 **3.10.6**, A 的前主子式是正的等价于 A 是正定的. □

4.6.9 定理 设 $A, B \in \mathbb{R}^{n\times n}$ 是 M-矩阵,则 AB 为 M-矩阵的充分必要条件是 $AB \in Z_n$.特别,若 $A, B \in \mathbb{R}^{2\times 2}$ 是 M-矩阵,则 AB 为 M-矩阵.

证明很容易,留作练习.

M-矩阵之和不一定是 M-矩阵.

例 设

$$A = \begin{bmatrix} 0.5 & -1 \\ 0 & 0.5 \end{bmatrix}, \quad B = \begin{bmatrix} 0.5 & 0 \\ -1 & 0.5 \end{bmatrix},$$

容易验证 A 和 B 都是 M-矩阵,但 $A+B$ 是奇异的.

4.6.10 定理 设 $A,B\in\mathbb{R}^{n\times n}$ 是 M-矩阵,而且 $B^{-1}A\in Z_n$,则 $\alpha A+(1-\alpha)B$ 对所有 $\alpha\in[0,1]$ 为 M-矩阵的充分必要条件是 $B^{-1}A$ 没有负的实特征值.

特别,如果 $B^{-1}A$ 是 M-矩阵,那么 $\alpha A+(1-\alpha)B$ 对所有 $\alpha\in[0,1]$ 为 M-矩阵.

证 显然,$\alpha A+(1-\alpha)B\in Z_n,\forall\alpha\in[0,1]$.

注意到

$$\alpha A+(1-\alpha)B=BC(\alpha),\quad C(\alpha)\equiv\left(\alpha B^{-1}A+(1-\alpha)I\right),$$

依 **4.6.9**,只要 $C(\alpha)$ 是 M-矩阵,则 $\alpha A+(1-\alpha)B$ 必为 M-矩阵.

设 λ 是 $B^{-1}A$ 的任一特征值,且 x 是相应的特征向量,则

$$C(\alpha)x=\left[\alpha B^{-1}A+(1-\alpha)I\right]x=(\alpha\lambda+1-\alpha)x,$$

表明 $\alpha\lambda+1-\alpha$ 是 $C(\alpha)$ 的特征值.

依 **4.6.5** 的(5),$C(\alpha)$ 是 M-矩阵等价于其每个实特征值是正的;也就是说,如果 $\alpha\lambda+1-\alpha\in\mathbb{R}$,则 $\alpha\lambda+1-\alpha>0$,即

$$\alpha(1-\lambda)<1.$$

显然,此不等式当 $\lambda>0$ 时对所有 $\alpha\in[0,1]$ 成立;而当 $\lambda<0$ 时 $\dfrac{1}{1-\lambda}<1$,则只对 $\alpha\in\left[0,\dfrac{1}{1-\lambda}\right)$ 成立. □

4.6.11 定理 设 $A,B\in Z_n$,而且 A 是 M-矩阵,$B\geq A$.则 $B^{-1}A$ 和 AB^{-1} 都是 M-矩阵,

$$B^{-1}A\leq I,\quad AB^{-1}\leq I.$$

而且对所有 $\alpha\in[0,1]$,$\alpha A+(1-\alpha)B$ 为 M-矩阵.

证 依 **4.6.6** 的(1),B 是 M-矩阵,$B^{-1}\geq 0$.令

$$R\equiv B-A,$$

则

$$R \geq 0,\ B^{-1}A = I - B^{-1}R \in Z_n,\ \ B^{-1}A \leq I.$$

因为 A 是 M-矩阵, 依 **4.6.5** 的(14),存在 $x>0$,使得 $Ax>0$,从而 $B^{-1}Ax>0$.因此,仍依 **4.6.5**, $B^{-1}A$ 是 M-矩阵.

同样,可证 $AB^{-1} \leq I$, AB^{-1} 是 M-矩阵.

最后,应用 **4.6.10**,即知对所有 $\alpha \in [0,1]$, $\alpha A + (1-\alpha)B$ 为 M-矩阵. □

4.6.12 定理 设 $A, B \in \mathbb{R}^{n\times n}$ 是 M-矩阵,则 $B^{-1}A$ 为 M-矩阵的充分必要条件是 $B^{-1}A \in Z_n$.

证 必要性显然.

反之,设 $B^{-1}A \in Z_n$,则从 **4.6.11** 的证明得知,存在 $x>0$ 使得 $B^{-1}Ax>0$.因此,依 **4.6.5**, $B^{-1}A$ 是 M-矩阵. □

4.6.13 定理 设 $A=[a_{ij}] \in \mathbb{R}^{n\times n}$ 是 M-矩阵,则至少对一个指标 i 成立

$$a_{ii} > \sum_{j=1,\, j\neq i}^{n} |a_{ij}|.$$

证 依 **4.6.5** 的(15),存在正对角矩阵

$$D = \operatorname{diag}(d_1, \cdots, d_n),$$

使得 AD 行严格对角优势.令

$$d_i = \min\{d_1, \cdots, d_n\},$$

则有

$$d_i a_{ii} > \sum_{j=1,\, j\neq i}^{n} d_j |a_{ij}| \geq d_i \sum_{j=1,\, j\neq i}^{n} |a_{ij}|. \qquad \square$$

4.6.14 定义 如果 $A \in Z_n$ 是对称的非奇异 M-矩阵,则称 A 为 **Stieltjes 矩阵**.

从 **4.6.8** 知, Stieltjes 矩阵也可定义为 Z_n 中的对称正定矩阵.

4.6.15 定义 如果 $A \in Z_n$ 是半正稳定而非正稳定的,则称 A 为**奇异 M-矩阵**.

通常,讲 M-矩阵是指非奇异 M-矩阵.从定义看,奇异 M-矩阵指 Z_n 中那些特征值有非负实部的矩阵,而其本身可以是奇异矩阵,也可以是非奇异矩阵.如此定义相对于 **4.6.4** 的 M-矩阵定义之三来说是很自然的.而且,基于"若 A 是奇异 M-矩阵,则对任意的 $\varepsilon > 0$, $A + \varepsilon I$ 是非奇异 M-矩阵"这一事实,容易推出奇异 M-矩阵的许多平行于 **4.6.5** 的等价条件.下面列举几个等价条件.

4.6.16 定理 对于 $A \in Z_n$,如下各条件是等价的:

(1) A 是半正稳定的.

(2) $A = \alpha I - P$, $P \ge 0$, $\alpha \ge \rho(P)$.

(3) A 的每个实特征值是非负的.

(4) A 的所有主子式是非负的.

(5) A 的所有 k 阶主子式之和是非负的, $k = 1, \cdots, n$.

证明留作练习.

4.6.17 定义 非奇异矩阵 $A \in \mathbb{R}^{n \times n}$ 称为**逆 M-矩阵**(inverse M-matrix),如果 A^{-1} 是 M-矩阵.

逆 M-矩阵从 M-矩阵那里承袭了大量的重要结构.

4.6.18 定理 设 $A \in \mathbb{R}^{n \times n}$ 是逆 M-矩阵,则

(1) $A \ge 0$.

(2) A 是正稳定的.

(3) 对 A 存在正对角 Lyapunov 解.

(4) 对任何正对角矩阵 D, DA 是逆 M-矩阵.

(5) A 的每个主子式是正的.

(6) A 的每个主子矩阵是逆 M-矩阵.

(7) 对任何正对角矩阵 D, $A+D$ 是逆 M-矩阵.

证 依 **4.6.17** 和 **4.6.4**, A^{-1} 是 M-矩阵,亦即 $A^{-1}\in Z_n$ 而且 $\left(A^{-1}\right)^{-1}\geq 0$.因此,以及由 **4.6.5** 有

(1) $A=\left(A^{-1}\right)^{-1}\geq 0$.

(2) A^{-1} 是正稳定的.依 **3.14.4**, $A=\left(A^{-1}\right)^{-1}$ 也是正稳定的.

(3) 存在正对角矩阵 D,使得 $DA^{-1}+\left(A^{-1}\right)^{\mathrm{T}}D$ 是正定矩阵.用 A 作合同变换,得

$$A^{\mathrm{T}}\left(DA^{-1}+\left(A^{-1}\right)^{\mathrm{T}}D\right)A=DA+A^{\mathrm{T}}D$$

仍是正定矩阵,这表明 D 也是 A 的 Lyapunov 解.

(4) 显然, $(DA)^{-1}=A^{-1}D^{-1}\in Z_n$,而且

$$\left((DA)^{-1}\right)^{-1}=DA\geq 0$$

(5),(6),(7)的证明从略,可作练习. □

4.6.19 定义 设 $A\in\mathbb{R}^{n\times n}$.

若 A 的所有主子式是正的,则称 A 为 **P-矩阵**;

若 A 的所有主子式是非负的,则称 A 为 **$\mathbf{P}_0$-矩阵**;

若 A 是 P_0-矩阵,而且对每个 k,至少有一个 k 阶主子式是正的, $k=1,\cdots,n$,则称 A 为 **$\mathbf{P}_0^+$-矩阵**.

M-矩阵自然地引伸出若干概念,包括 P-矩阵,D-稳定矩阵等,它们在其它一些范围中也是有意义的.

4.6.20 定理 设 $A=[a_{ij}]\in\mathbb{R}^{n\times n}$ 是 P-矩阵, $D\in\mathbb{R}^{n\times n}$ 是非负对角矩阵,则

$$\det(A+D)\geq\det A>0, \tag{6.9}$$

其中等号仅当 $D=0$ 时成立.

设 A 是 P_0^+-矩阵, $t\in\mathbb{R}, t>0$, 则

$$\det(A+tI)>\det A>0. \tag{6.10}$$

由此, P_0^+-矩阵的实特征值必是正的.

证 设

$$D=\operatorname{diag}(d_1,\cdots,d_n),$$

并且

$$D_i\equiv\operatorname{diag}(0,\cdots,0,d_i,0,\cdots,0)\in\mathbb{R}^{n\times n},\quad i=1,\cdots,n$$

利用行列式的 Laplace 展开,有

$$\begin{aligned}\det(A+D_1)&=(-1)^{1+1}(a_{11}+d_1)\det A_{11}\\&\quad+(-1)^{2+1}a_{21}\det A_{21}+\cdots\\&\quad+(-1)^{n+1}a_{n1}\det A_{n1}\\&=\det A+d_1\det A_{11}.\end{aligned} \tag{6.11}$$

于是,因 A 是 P-矩阵,当 $d_1>0$ 时有

$$\det(A+D_1)>\det A>0,$$

显然, $A+D_1$ 仍然是 P-矩阵.

重复如此论证,最后可得

$$\begin{aligned}0&<\det A\le\det(A+D_1)\\&\le\det(A+D_1+D_2)\le\cdots\\&\le\det(A+D_1+\cdots+D_n)=\det(A+D).\end{aligned} \tag{6.12}$$

而且从推证中得知,只要有某个 $d_i\neq0$,(6.9)便成立严格不等号.

现在设

$$d_1=\cdots=d_n=t.$$

根据 P_0^+-矩阵的定义,不妨假设 $\det A_{11}>0$,于是当 $t>0$ 时,由(6.11)和(6.12)知(6.10)成立. □

如同 M-矩阵那样, P-矩阵有许多不同的等价条件.

4.6.21 定理 设 $A=[a_{ij}]\in\mathbb{R}^{n\times n}$,下列各条件等价:

(1) A 的所有主子式是正的,即 A 是 P-矩阵.

(2) 对于每个非零的 $x=(x_1,\cdots,x_n)^{\mathrm{T}}\in\mathbb{R}^n$,Hadamard 积 $x\circ(Ax)$ 的某个元素是正的,也就是说,存在某 $k\in\{1,\cdots,n\}$,使得

$$x_k(Ax)_k=x_k(a_{k1}x_1+\cdots+a_{kn}x_n)>0.$$

(3) 对于每个非零的 $x=(x_1,\cdots,x_n)^{\mathrm{T}}\in\mathbb{R}^n$,存在正对角矩阵 $D=D(x)\in\mathbb{R}^{n\times n}$,使得 $x^{\mathrm{T}}(DA)x>0$.

(4) 对于每个非零的 $x=(x_1,\cdots,x_n)^{\mathrm{T}}\in\mathbb{R}^n$,存在非负对角矩阵 $\Lambda=\Lambda(x)\in\mathbb{R}^{n\times n}$,使得 $x^{\mathrm{T}}(\Lambda A)x>0$.

(5) A 的每个主子矩阵的每个实特征值是正的.

证 (1)⇒(2): 设 $x=(x_1,\cdots,x_n)^{\mathrm{T}}\in\mathbb{R}^n$, $x\neq 0$.用

$$\hat{x}\equiv(x_{j_1},\cdots,x_{j_k})^{\mathrm{T}}\in\mathbb{R}^k$$

表示由 x 的所有非零分量组成的向量,其中

$$1\le j_1<\cdots<j_k\le n,\ 1\le k\le n.$$

同时考虑主子矩阵

$$\hat{A}\equiv A(\{j_1,\cdots,j_k\}),$$

注意

$$(\hat{A}\hat{x})_i=(Ax)_{j_i},\ i=1,\cdots,k.$$

假定 $x\circ(Ax)\le 0$,则 $\hat{x}\circ(\hat{A}\hat{x})\le 0$,亦即

$$x_{j_i}(Ax)_{j_i}\equiv x_{j_i}(a_{j_ij_1}x_{j_1}+\cdots+a_{j_ij_k})\le 0,\ i=1,\cdots,k.$$

由此,

$$(Ax)_{j_i}/x_{j_i}\le 0,\ i=1,\cdots,k.$$

令

$$d_i\equiv-\frac{(Ax)_{j_i}}{x_{j_i}}\ge 0,\ i=1,\cdots,k,$$

并记 $D \equiv \operatorname{diag}(d_1, \cdots, d_k)$，$D$ 是非负对角矩阵.于是

$$\hat{A}\hat{x} = -D\hat{x} \quad 或 \quad (\hat{A} + D)\hat{x} = 0 .$$

但这是不可能的,因为依 **4.6.20**,

$$\det(\hat{A} + D) \geq \det \hat{A} > 0 ,$$

而且 $\hat{A} + D$ 是非奇异的.因此 $x \circ (Ax)$ 有某元素是正的.

(2)$\Rightarrow$(3): 设

$$x = (x_1, \cdots, x_n)^{\mathrm{T}} \in \mathbb{R}^n, \quad x \neq 0 ,$$

而且指标 $k \in \{1, \cdots, n\}$ 使得 $x_k(Ax)_k > 0$，则存在 $\varepsilon > 0$ 成立

$$x_k(Ax)_k + \varepsilon \sum_{i=1, i \neq k}^{n} x_i(Ax)_i > 0 .$$

于是,可取 $D \equiv \operatorname{diag}(d_1, \cdots, d_n)$，其中

$$d_k = 1 ; \quad d_i = \varepsilon, \ \forall i \neq k .$$

(3)$\Rightarrow$(4): 显然.

(4)$\Rightarrow$(5): 设

$$\hat{A} \equiv A(\{j_1, \cdots, j_k\})$$

是 A 的主子矩阵，$1 \leq j_1 < \cdots < j_k \leq n$，$1 \leq k \leq n$；又设 $\lambda \in \mathbb{R}$ 是 $\hat{A}$ 的特征值,对应的特征向量是

$$\xi = (\xi_1, \cdots, \xi_k)^{\mathrm{T}} \in \mathbb{R}^k, \quad \xi \neq 0 .$$

取

$$x = (x_1, \cdots, x_n)^{\mathrm{T}} \in \mathbb{R}^n ,$$

其中

$$x_{j_i} = \xi_i, \ i = 1, \cdots, k ; \qquad x_i = 0, \ \forall i \notin \{j_1, \cdots, j_k\} .$$

现设 $\Lambda = \Lambda(x) \in \mathbb{R}^{n \times n}$ 是非负对角矩阵，使得 $x^{\mathrm{T}}(\Lambda A)x > 0$.
又设 $\hat{\Lambda} \equiv \Lambda(\{j_1, \cdots, j_k\})$ 是 Λ 的主子矩阵.于是

$$0 < x^{\mathrm{T}}(\Lambda A)x = (\Lambda x)^{\mathrm{T}}(Ax) = (\hat{\Lambda})^{\mathrm{T}}(\lambda\xi) = \lambda(\xi^{\mathrm{T}}\hat{\Lambda}\xi) ,$$

这蕴涵 λ 和 $\xi^{\mathrm{T}}\Lambda\xi$ 都是正的.

(5)$\Rightarrow$(1): A 是实的,它的复特征值必共轭成对出现,其积是正的;又根据(5)所设,A 的实特征值全是正的.而 $\det A$ 是所有特征值(按代数重数计)之积,所以 $\det A>0$.这样的论证适用于 A 的每一个主子矩阵. □

4.6.22 定义 $A\in\mathbb{C}^{n\times n}$ 称为 **D-稳定矩阵**,如果对所有正对角矩阵 $D\in\mathbb{R}^{n\times n}$, DA 是正稳定的.

$A\in\mathbb{C}^{n\times n}$ 称为 **D-半稳定矩阵**,如果对所有正对角矩阵 $D\in\mathbb{R}^{n\times n}$, DA 是半正稳定的.

4.6.23 定理 (1) 每个 M-矩阵是 P-矩阵和 D-稳定矩阵.

(2) 若 $A\in\mathbb{C}^{n\times n}$ 是正定的,或者更一般地,若 $A+A^*$ 是正定的,则 A 是 D-稳定矩阵.

(3) D-稳定矩阵是非奇异的,而且其逆也是 D-稳定矩阵.

(4) D-稳定矩阵的主子矩阵是 D-半稳定的,而不一定是 D-稳定的.

证 (1),(2),(3)的证明留作练习.

现在证明(4).

设 $A\in\mathbb{C}^{n\times n}$ 是 D-稳定的; $\alpha\subset\{1,\cdots,n\}$ 是一指标集.

设 $\Lambda\in\mathbb{R}^{n\times n}$ 是对角矩阵,其对角元素属 α 指明的位置上的为 1,其余位置上的为 $\varepsilon>0$.

考虑 $D\Lambda A$,其中 $D\in\mathbb{R}^{n\times n}$ 是任一正对角矩阵.于是,一方面,由于 A 是 D-稳定的,根据定义,$D\Lambda A$ 是正稳定的,即其所有特征值具有正实部;另一方面,只要 $\varepsilon>0$ 充分小,$\lambda(D\Lambda A)$ 可以近似包含 $\lambda(D(\alpha)A(\alpha))$ 到任意程度.因此主子矩阵 $A(\alpha)$ 至少是 D-半稳定的.

再来看一个例子,说明D-稳定矩阵的主子矩阵可以不是D-稳定的.设

$$A=\begin{bmatrix}0 & -1\\ 1 & 1\end{bmatrix},\quad D=\begin{bmatrix}a & 0\\ 0 & b\end{bmatrix},\quad a>0,b>0.$$

DA 的特征方程是

$$\det(\lambda I - DA) = \lambda^2 - b\lambda + ab = 0,$$

由此得

$$\lambda(DA) = \left\{ \frac{b + \sqrt{b^2 - 4ab}}{2}, \frac{b - \sqrt{b^2 - 4ab}}{2} \right\},$$

DA 的特征值当 $b^2 - 4ab \geq 0$ 时是两个正实数,当 $b^2 - 4ab < 0$ 时是一对具有正实部的共轭复数.因此,A 是正稳定的,但其主子矩阵 $A(\{1\}) = [0]$ 只是 D-半稳定的,而不是 D-稳定的. □

4.6.24 定理 设 $A \in \mathbb{R}^{n\times n}$.

(1) 若 A 是 D-稳定的,则 A 是 P_0^+-矩阵.

(2) 若 A 有正对角 Lyapunov 解,则 A 是 D-稳定的.

证 设 A 是 D-稳定矩阵.考虑任一 k 阶主子矩阵 $A(\alpha)$,其中 $\alpha \subset \{1,\cdots,n\}$ 是 k 个指标的集合;依 **4.6.23** 的(4),k 阶主子式 $\det A(\alpha) \geq 0$.这表明 A 是 P_0-矩阵.

注意到从 A 的特征多项式 $\det(\lambda I - A)$ 的根与系数的关系,显示 A 的所有 k 阶主子式之和 $E_k(A)$ 与 A 的特征值 $\lambda_1,\cdots,\lambda_n$ 之间的关系式

$$E_k(A) = \sum_{1\leq j_1<\cdots<j_k\leq n} \lambda_{j_1}\cdots\lambda_{j_k},$$

$\pm E_k(A)$ 是特征多项式中 λ^{n-k} 的系数;再注意到 A 是实的和正稳定的;从而不可能所有 k 阶主子式都等于零,即至少有一个 k 阶主子式是正的.因此,A 是 P_0^+-矩阵.(1)得证.

现在假定存在正对角矩阵 $\Lambda \in \mathbb{R}^{n\times n}$,使得 $\Lambda A + A^T\Lambda \equiv B$ 是正定的.设 $D \in \mathbb{R}^{n\times n}$ 是任一正对角矩阵,考虑 DA,注意到

$$(\Lambda D^{-1})(DA) + (DA)^{\mathrm{T}}(\Lambda D^{-1}) = \Lambda A + A^{\mathrm{T}}\Lambda = B,$$

所以正定矩阵 ΛD^{-1} 是 DA 的 Lyapunov 解,DA 是正稳定的.因

此,A 是 D-稳定的. □

这个定理给出 A 是 D-稳定矩阵的一个基本的必要条件和一个基本的充分条件.

4.6.25 定义 设 $A \in Z_n$.记

$$\ell(A) \equiv \min\{\mathrm{Re}(\lambda) : \lambda \in \lambda(A)\}, \tag{6.13}$$

$\ell(A)$称为 Z-矩阵 A 的**最小特征值**.

4.6.26 定理 设 $A = [a_{ij}] \in Z_n$,则

(1) $\ell(A) \in \lambda(A)$.

(2) $\ell(A) \ge \min\limits_{1\le i\le n} \sum\limits_{j=1}^{n} a_{ij}$

(3) 如果存在向量 $y \in \mathbb{R}^n$ 使得 $Ay \ge \beta y$,那么 $\ell(A) \ge \beta$.

(4) 若 A 不可约,则在(3)中可将 y 是正的减弱为非零非负的,便可保证 $\ell(A) \ge \beta$.

证 (1) 将 A 表示成

$$A = \alpha I - P,\ \ \alpha \in \mathbb{R},\ \ P \ge 0. \tag{6.14}$$

依 **4.6.2**,$\alpha - \rho(P)$是 A 的特征值,且

$$\mathrm{Re}\,\lambda \ge \alpha - \rho(P),\quad \forall \lambda \in \lambda(A), \tag{6.15}$$

因此 $\ell(A) = \alpha - \rho(P) \in \lambda(A)$.

(2) 利用表示式(6.14).设 A 不可约,则 P 也不可约.依 **4.4.2**,存在 P 的属于 $\rho(P)$的非负的特征向量 $x = (x_1, \cdots, x_n)^{\mathrm{T}}$,

$$Px = \rho(P)x.$$

注意到 $\ell(A) = \alpha - \rho(A)$,

$$Ax = (\alpha I - P)x = \ell(A)x. \tag{6.16}$$

由此,并令

$$x_{i_0} = \max\{x_1, \cdots, x_n\},$$

有

$$\ell(A)=\frac{1}{x_{i_0}}\left(a_{i_0 1}x_1+\cdots+a_{i_0 n}x_n\right)=\sum_{j=1,j\neq i_0}^{n}a_{i_0 j}\frac{x_j}{x_{i_0}}+a_{i_0 i_0}.$$

再注意到

$$\frac{x_j}{x_{i_0}}\le 1,\quad a_{i_0 j}\le 0,\quad j=1,\cdots,n,\quad j\neq i_0,$$

于是

$$\ell(A)\ge\sum_{j=1}^{n}a_{i_0 j}\ge\min_{1\le i\le n}\sum_{j=1}^{n}a_{ij}.$$

A 可约情形的证明留作练习.

(3) 显然,对于 A^{T},因有与 A 完全相同的特征值,故同样可得形如(6.16)的关系式:

$$A^{\mathrm{T}}x=\ell\left(A^{\mathrm{T}}\right)x=\ell(A)x,$$

其中 $x\ge 0$.再利用 $Ay\ge\beta y$,得

$$\ell(A)x^{\mathrm{T}}y=x^{\mathrm{T}}Ay\ge\beta x^{\mathrm{T}}y. \tag{6.17}$$

由于 $x^{\mathrm{T}}y>0$,因此 $\ell(A)\ge\beta$.

(4) 仍利用表示式(6.14).因 A 不可约,P 必不可约,依 **4.2.4**,存在 P 的属于 $\rho(P)$ 的特征向量 $x>0$;故只要 y 是非零非负的便有 $x^{\mathrm{T}}y>0$,从而由(6.17)得 $\ell(A)\ge\beta$. □

4.6.27 定理 设 $A\in\mathbb{R}^{n\times n}$ 是M-矩阵,则

(1) 如果将 A 表示为

$$A=\alpha I-P,\ \alpha\in\mathbb{R},\ P\ge 0,$$

那么

$$\lambda(A)\subset\left\{z\in\mathbb{C}:|z-\alpha|\le\rho(P)\right\}. \tag{6.18}$$

而且

$$\ell(A)=1/\rho\left(A^{-1}\right)=\alpha-\rho(P). \tag{6.19}$$

(2) $\det A\ge[\ell(A)]^n$.

证 依 **4.6.2**,已有(6.18).

依 **4.6.26** 的(1),已得

$$\ell(A)=\alpha-\rho(P).$$

因 $A^{-1}\geq 0$, $\rho\left(A^{-1}\right)$是其模最大的特征值,故$1/\rho\left(A^{-1}\right)$是 A 的模最小的特征值.推出

$$\ell(A)=1/\rho\left(A^{-1}\right).$$

依 **4.6.5** 的(8), $\det A>0$.因此,若 $\lambda(A)=\{\lambda_1,\cdots,\lambda_n\}$,则

$$\det A=\prod_{i=1}^{n}\lambda_i=\prod_{i=1}^{n}|\lambda_i|\geq\prod_{i=1}^{n}(\operatorname{Re}\lambda_i)\geq[\ell(A)]^n. \qquad \square$$

4.6.28 定理 设 $A,B\in\mathbb{R}^{n\times n}$ 是 M-矩阵, $A\geq B$,则

$$\ell(A)\geq\ell(B).$$

证 依 **4.6.6**, $B^{-1}\geq A^{-1}\geq 0$.先假定 B^{-1} 不可约,依 **4.2.4**,存在左特征向量 $y^{\mathrm{T}}>0$,使得

$$y^{\mathrm{T}}B^{-1}=\rho\left(B^{-1}\right)y^{\mathrm{T}}.$$

对于 A^{-1},依 **4.4.2**,存在右特征向量 $x\geq 0$,使得

$$A^{-1}x=\rho\left(A^{-1}\right)x,$$

于是

$$\rho\left(B^{-1}\right)y^{\mathrm{T}}x=y^{\mathrm{T}}B^{-1}x\geq y^{\mathrm{T}}A^{-1}x=\rho\left(A^{-1}\right)y^{\mathrm{T}}x.$$

由于 $y^{\mathrm{T}}x\neq 0$,因此

$$\ell(B)=1/\rho\left(B^{-1}\right)\leq 1/\rho\left(A^{-1}\right)=\ell(A).$$

对于 B^{-1} 可约的情形,可用 $\varepsilon>0$ 填补其零元素,便可利用如上结论,然后让 $\varepsilon\to 0$ 取极限,同样可得 $\ell(B)\leq\ell(A)$. $\square$

4.6.29 定理 设 $A\in\mathbb{R}^{n\times n}$ 是 M-矩阵, $D=\operatorname{diag}(d_1,\cdots,d_n)$是正对角矩阵,则

$$\ell(DA)\geq\ell(A)\min_{1\leq i\leq n}d_i. \tag{6.20}$$

证 注意到 DA 是 M-矩阵,

$$(DA)^{-1} = A^{-1}D^{-1},\ A^{-1} \geq 0,\ D^{-1} > 0,$$

从而

$$\ell(DA) = \rho\left(A^{-1}D^{-1}\right)^{-1} \geq \rho\left(A^{-1}\left(\max_{1\leq i\leq n} d_i^{-1}\right)I\right)^{-1}$$

$$= \left(\rho\left(A^{-1}\right)\max_{1\leq i\leq n} d_i^{-1}\right)^{-1} = \ell(A)\min_{1\leq i\leq n} d_i. \qquad \square$$

4.6.30 定理 设 $A = [a_{ij}] \in \mathbb{R}^{n\times n}$ 是行严格对角优势的 L-矩阵, $B = [b_{ij}] \in \mathbb{R}^{n\times n}, B \geq 0$,则

$$\rho\left(A^{-1}B\right) \geq \min_{1\leq i\leq n}\left(\sum_{j=1}^{n} b_{ij} \Big/ \sum_{j=1}^{n} a_{ij}\right). \qquad (6.21)$$

证 令

$$\xi = \left(\xi_1, \cdots, \xi_n\right)^{\mathrm{T}} \equiv A^{-1}Be,$$

这里 $e \equiv (1, \cdots, 1)^{\mathrm{T}} \in \mathbb{R}^n$.由假定并依 **2.2.6**,知 $\xi \geq 0$.设

$$\xi_{i_0} = \min_{1\leq i\leq n} \xi_i,$$

从 $A\xi = Be$ 及 A 还是 L-矩阵,得

$$\sum_{j=1}^{n} b_{i_0 j} = \sum_{j=1}^{n} a_{i_0 j}\xi_j \leq \xi_{i_0}\sum_{j=1}^{n} a_{i_0 j}. \qquad (6.22)$$

记

$$\xi_{ij} = e_i^{\mathrm{T}} A^{-1} B e_j.$$

依 **4.1.8**,有

$$\rho\left(A^{-1}B\right) \geq \min_{1\leq i\leq n}\sum_{j=1}^{n}\xi_{ij} = \min_{1\leq i\leq n}\xi_i = \xi_{i_0}, \qquad (6.23)$$

综合(6.22)和(6.23)即得(6.21). $\square$

4.7 H-矩阵

4.7.1 定义 设 $A=[a_{ij}]\in\mathbb{C}^{n\times n}$,并设 $\mathscr{M}(A)\equiv[m_{ij}]\in\mathbb{R}^{n\times n}$,

$$m_{ij}=\begin{cases}|a_{ii}|, & j=i,\\ -|a_{ij}|, & j\neq i,\end{cases}\quad i,j=1,\cdots,n, \tag{7.1}$$

$\mathscr{M}(A)$称为 A 的**比较矩阵**(comparison matrix).

4.7.2 定义 设 $A\in\mathbb{C}^{n\times n}$,如果 A 的比较矩阵 $\mathscr{M}(A)$是非奇异的 M-矩阵,则称 A 为**非奇异 H-矩阵**,简称 **H-矩阵**.

4.7.3 定理 设 $A\in\mathbb{C}^{n\times n}$,则有如下明显性质:

(1) $\mathscr{M}(A)\in Z_n$.

(2) $\mathscr{M}(A)=A$的充分必要条件是 $A\in Z_n$.

(3) $\mathscr{M}(A)$可表示为非负对角矩阵与具有零对角的非负矩阵之差:

$$\mathscr{M}(A)=|I\circ A|-(|A|-|I\circ A|), \tag{7.2}$$

其中

$$I\circ A=\operatorname{diag}(a_{11},\cdots,a_{nn})$$

是 I 与 A 的 Hadamard 积,而

$$|X|\equiv[\,|x_{ij}|\,]$$

表示 $X=[x_{ij}]\in\mathbb{C}^{n\times n}$ 的逐个元素取绝对值后的矩阵.

(4) A 为 M-矩阵的充分必要条件是

$$\mathscr{M}(A)=A,$$

而且 A 为 H-矩阵.

(5) 若 A 是 M-矩阵,则(7.2)成为

$$A=(I\circ A)-[(I\circ A)-A]. \tag{7.3}$$

此表示式可供替代非常有用的通常表示式 $A = \alpha I - P$.

证明留作练习.

4.7.4 定理 设 $A, B \in \mathbb{C}^{n\times n}$, A 是 M-矩阵, $\mathcal{M}(B) \geq A$, 则

(1) B 是 H-矩阵.

(2) B 是非奇异的.

(3) $A^{-1} \geq \left|B^{-1}\right| \geq 0$.

(4) $|\det B| \geq \det A > 0$.

证 这一定理是将 **4.6.6** 的基本结果推广至复矩阵 B.

(1)直接从 **4.6.6** 的(1)及 **4.7.2** 推出.

现在取对角酉矩阵

$$D \equiv \operatorname{diag}\left(\overline{b_{11}}/|b_{11}|, \cdots, \overline{b_{nn}}/|b_{nn}|\right),$$

则 DB 的主对角元素是正数

$$|b_{11}|, \quad \cdots, \quad |b_{nn}|.$$

依从 4.6.5 的(3),将 A 表示为

$$A = \alpha I - P, \quad P \geq 0, \quad \alpha > \rho(P).$$

并令

$$R \equiv \alpha I - DB.$$

由于

$$|R| = \begin{bmatrix} |\alpha - |b_{11}|| & |b_{12}| & \cdots & |b_{1n}| \\ |b_{21}| & |\alpha - |b_{22}|| & \cdots & |b_{2n}| \\ \cdots & \cdots & \cdots & \cdots \\ |b_{n1}| & |b_{n2}| & \cdots & |\alpha - |b_{nn}|| \end{bmatrix}$$

$$\leq P + A - \mathcal{M}(A) \leq P,$$

依 **4.1.4**,

$$\rho(R) \leq \rho(|R|) \leq \rho(P) < \alpha,$$

因此 $DB = \alpha I - R$ 可逆,故 B 必非奇异.而且

$$\left|B^{-1}\right|=\left|(DB)^{-1}\right|=\left|\alpha^{-1}\left(I-\alpha^{-1}R\right)^{-1}\right|$$

$$=\left|\sum_{k=0}^{\infty}\alpha^{-k-1}R^{k}\right|\leq\sum_{k=0}^{\infty}\frac{1}{\alpha^{k+1}}P^{k}=A^{-1}.$$

仿照 **4.6.6** 的(3)的推导,可得

$$\frac{\det A_{11}}{\det A}=\left(A^{-1}\right)_{nn}\geq\left|\left(B^{-1}\right)_{nn}\right|=\left|\frac{\det B_{11}}{\det B}\right|,$$

从而$|\det B|\geq\det A$. □

4.7.5 定理 严格对角优势矩阵和不可约弱严格对角优势矩阵是 H-矩阵.

证 设$A=[a_{ij}]\in\mathbb{C}^{n\times n}$是严格对角优势的或不可约弱严格对角优势的.依 **2.2.3**, $\mathcal{M}(A)\in Z_n$是非奇异的.任取

$$\alpha>\min_{1\leq i\leq n}|a_{ii}|,$$

则有

$$P=\alpha I-\mathcal{M}(A)\geq 0.$$

依 **4.2.4**, $\rho(P)$是P的特征值.于是,依 Gerschgorin 圆盘定理 **1.6.23**,存在i, $1\leq i\leq n$,使得

$$\rho(P)\leq\left|\alpha-|a_{ii}|\right|+\sum_{j=1,j\neq i}^{n}\left|a_{ij}\right|=\alpha-\left(|a_{ii}|-\sum_{j=1,j\neq i}^{n}\left|a_{ij}\right|\right)\leq\alpha.$$

注意到$\alpha-\rho(P)\geq 0$是$\mathcal{M}(A)$的特征值,而$\mathcal{M}(A)$是非奇异的,从而必有$\alpha>\rho(P)$.因此, $\mathcal{M}(A)=\alpha I-P$是 M-矩阵,故$A$是 H-矩阵. □

4.7.6 定义 对于任何矩阵$A\in\mathbb{C}^{n\times n}$,按照(6.13), $\ell(\mathcal{M}(A))$是有意义的,定义

$$\ell(A)\equiv\ell(\mathcal{M}(A)). \tag{7.4}$$

这是将 Z-矩阵的最小特征值的定义 **4.6.25** 推广至一般矩阵.

4.7.7 定理 $A \in \mathbb{C}^{n \times n}$ 为 H-矩阵的充分必要条件是 $\ell(A) > 0$.

这一定理很容易证明,留作练习.

4.8 完全非负矩阵简述

4.8.1 定义 $A \in \mathbb{R}^{n \times n}$ 称为**完全非负矩阵**(totally nonnegative matrix),如果 A 的所有各阶子式全是非负的.

$A \in \mathbb{R}^{n \times n}$ 称为**完全正矩阵**(totally positive matrix),如果 A 的所有各阶子式全是正的.

4.8.2 定理 设 $A \in \mathbb{R}^{n \times n}$.

(1) 若 A 是完全非负矩阵,则 $A \geq 0$,即 A 必为非负矩阵.

(2) 若 A 是完全正矩阵,则 $A > 0$,即 A 必为正矩阵.

(3) 若 A 是对称的完全正矩阵,则 A 是正定矩阵.

证明留作练习.

4.8.3 定义 设 $A \in \mathbb{R}^{n \times n}$ 是完全非负矩阵,而且存在正整数 k,使得 A^k 是完全正矩阵,则称 A 为**振荡矩阵**(oscillatory matrix).

振荡矩阵这一术语出于在振荡理论中的应用.

4.8.4 定理 设 $A \in \mathbb{R}^{n \times n}$ 是完全非负矩阵.则 A 为振荡矩阵的充分必要条件是 $\det A > 0$,而且在包含主对角的三对角带上的所有元素全为正的.

证明从略.

4.8.5 定理 设 $A \in \mathbb{R}^{n \times n}$ 是完全非负矩阵,则

(1) A 的所有特征值是实的和非负的.

(2) 如果 A 是振荡矩阵,那么 A 的所有特征值是正的而且是相异的.

证明从略.

5 标准形矩阵及其变换矩阵

5.1 Jordan 标准形和相似性

5.1.1 引言 设 $A \in \mathbb{C}^{n \times n}$.

若 A 有 n 个不同的特征值 $\lambda_1, \cdots, \lambda_n$,则根据 **1.6.5**, A 有 n 个线性无关的特征向量 $x^{(1)}, \cdots, x^{(n)}$.于是

$$AP = P\operatorname{diag}(\lambda_1, \cdots, \lambda_n), \quad P \equiv \begin{bmatrix} x^{(1)} & \cdots & x^{(n)} \end{bmatrix}. \tag{1.1}$$

由此

$$P^{-1}AP = \operatorname{diag}(\lambda_1, \cdots, \lambda_n). \tag{1.2}$$

这就是说,对于 A 没有重特征值的情形, A 相似于以其特征值为对角元素的对角矩阵,或者说 A 可以通过相似变换约化为对角矩阵.

然而,对于 A 有一个或多个重特征值的情形,问题远非那么简单.

定义矩阵 $J_k(\lambda) \in \mathbb{C}^{k \times k}$ 如下:

$$J_1(\lambda) = [\lambda];$$

$$J_k(\lambda) = \begin{bmatrix} \lambda & 1 & 0 & \cdots & 0 \\ 0 & \ddots & \ddots & \ddots & \vdots \\ \vdots & \ddots & \ddots & \ddots & 0 \\ \vdots & & \ddots & \lambda & 1 \\ 0 & \cdots & \cdots & 0 & \lambda \end{bmatrix}, \quad k > 1. \tag{1.3}$$

容易证明:如此矩阵只有一个特征值 λ,其代数重数为 k;其几何重数则为1,即只有一个线性无关的特征向量,可取为

$$x = e_1 \equiv (1,0,\cdots,0)^{\mathrm{T}} \in \mathbb{C}^k,$$

而且不能通过相似变换将矩阵对角化.

一般地说,一个 $n \times n$ 矩阵 A 通过相似变换能约化成怎样的最紧形式呢? 答案包含在下述著名定理中.

5.1.2 定理 设 $A \in \mathbb{C}^{n\times n}$ 有 $r \le n$ 个不同的特征值

$$\lambda_1,\cdots,\lambda_r,$$

代数重数分别为

$$n_1,\cdots,n_r, \quad n_1+\cdots+n_r=n,$$

则存在非奇异 $P \in \mathbb{C}^{n\times n}$,使得

$$P^{-1}AP = J \equiv \mathrm{diag}(J_1,\cdots,J_r), \tag{1.4}$$

其中

$$J_i = \mathrm{diag}\left(J_{n_i(1)}(\lambda_i),\cdots,J_{n_i(k_i)}(\lambda_i)\right),$$

$$\sum_{k=1}^{k_i} n_i(k) = n_i, \quad i=1,\cdots,r, \tag{1.5}$$

$J_{n_i(k)}(\lambda_i)$ 是由(1.3)定义的 $n_i(k)\times n_i(k)$ 矩阵.

在不考虑因 P 的不同选取而引起诸 J_i 顺序改变和诸 $J_{n_i(k)}$ 顺序改变的意义下,J 是唯一的.

此定理证明见 **9.6.22**.

5.1.3 定义 形如(1.3)的矩阵 $J_k(\lambda)$ 称为 **Jordan 块**.形如(1.4)的矩阵 J 称为矩阵 A 的 **Jordan 标准形**.并且常称 J 为其所含的各 Jordan 块的**直和**.

注: Jordan 块也可以取下三角矩阵形式,即不在上次主对角线取1,而在下次主对角线取1.

5.1.4 定义　如果$A \in \mathbb{C}^{n\times n}$的 Jordan 标准形的所有 Jordan 块都是1阶的,则称A为**非亏损矩阵**(non-defective matrix).;否则,称A为**亏损矩阵**(defective matrix).

A为$n\times n$亏损矩阵的直观含义是:A的线性无关的特征向量少于n个.

A为$n\times n$非亏损矩阵,即其Jordan标准形J是以A的特征值为对角元素的对角矩阵:

$$J = \operatorname{diag}(\lambda_1, \cdots, \lambda_n). \tag{1.6}$$

如此,通常也称非亏损矩阵为**可对角化矩阵**(diagonalizable matrix),或称为**(简)单矩阵**(simple matrix).

5.1.5 定理　设$A \in \mathbb{C}^{n\times n}$,则下列三个条件等价:

(1) A是非亏损矩阵.

(2) 成立

$$A = Q\operatorname{diag}(\lambda_1, \cdots, \lambda_n)Q^{-1}, \tag{1.7}$$

其中$\lambda_1, \cdots, \lambda_n$是$A$的特征值,$Q$是以相应特征向量为列的非退化矩阵.

(3) A的每个特征值的几何重数等于其代数重数.

此定理证明比较简单,留作练习.

5.1.6 定义　如果$A \in \mathbb{C}^{n\times n}$的每个特征值的几何重数均为1,则称$A$为**非减次矩阵**(non-derogatory matrix);否则,称A为**减次矩阵**(derogatory matrix).

根据定义,A是非减次矩阵,即其每个特征值都只有1个线性无关的特征向量;也就是说,在A的Jordan标准形中每个特征值只对应1个Jordan块.

5.1.7 定义　两个可对角化矩阵$A, B \in \mathbb{C}^{n\times n}$称为**同时可对角化**的,

如果存在一个相似矩阵 $S \in \mathbb{C}^{n\times n}$,使得

$$S^{-1}AS \quad 和 \quad S^{-1}BS$$

同为对角矩阵.

5.1.8 引理　设 $A \in \mathbb{C}^{m\times m}, B \in \mathbb{C}^{n\times n}$,并设

$$C = \begin{bmatrix} A & 0 \\ 0 & B \end{bmatrix}$$

是 A 和 B 的直和,则 C 可对角化的充分必要条件是 A 和 B 两者可对角化.

证　如果存在非奇异矩阵 $S_1 \in \mathbb{C}^{m\times m}$ 和 $S_2 \in \mathbb{C}^{n\times n}$,使得 $S_1^{-1}AS_1$ 和 $S_2^{-1}BS_2$ 是对角矩阵,那么取 S 是 S_1 和 S_2 的直和

$$S = \begin{bmatrix} S_1 & 0 \\ 0 & S_2 \end{bmatrix},$$

便有 $S^{-1}CS$ 是对角矩阵.

反之,如果 C 是可对角化的,

$$S^{-1}CS = \Lambda \equiv \operatorname{diag}(\lambda_1, \lambda_2, \cdots, \lambda_{m+n}),$$

其中 $S \in \mathbb{C}^{(m+n)\times(m+n)}$ 是非奇异矩阵.将 S 写成

$$S = \begin{bmatrix} s^{(1)} & s^{(2)} & \cdots & s^{(m+n)} \end{bmatrix},$$

其中

$$s^{(i)} = \begin{bmatrix} \xi^{(i)} \\ \eta^{(i)} \end{bmatrix} \in \mathbb{C}^{m+n}, \quad \xi^{(i)} \in \mathbb{C}^m, \eta^{(i)} \in \mathbb{C}^n,$$

$$i = 1,2,\cdots,m+n.$$

那么,从 $Cs^{(i)} = \lambda_i s^{(i)}, i = 1,2,\cdots,m+n$,推出

$$A\xi^{(i)} = \lambda_i \xi^{(i)}, \quad B\eta^{(i)} = \lambda_i \eta^{(i)}, \quad i = 1,2,\cdots,m+n.$$

现在考虑向量集合

$$\left\{\xi^{(1)},\xi^{(2)},\cdots,\xi^{(m+n)}\right\}.$$

若其线性无关向量个数少于 m,则矩阵

$$\left[\xi^{(1)} \quad \xi^{(2)} \quad \cdots \quad \xi^{(m+n)}\right]\in\mathbb{C}^{m\times(m+n)}$$

的列秩(并因而行秩)小于 m.同样,若集合 $\left\{\xi^{(1)},\xi^{(2)},\cdots,\xi^{(m+n)}\right\}$ 线性无关向量个数少于 n,则矩阵

$$\left[\eta^{(1)} \quad \eta^{(2)} \quad \cdots \quad \eta^{(m+n)}\right]\in\mathbb{C}^{n\times(m+n)}$$

的列秩(并因而行秩)小于 n.由此可知,以上两种情形,无论哪一种,矩阵

$$S=\begin{bmatrix}\xi^{(1)} & \xi^{(2)} & \cdots & \xi^{(m+n)}\\ \eta^{(1)} & \eta^{(2)} & \cdots & \eta^{(m+n)}\end{bmatrix}$$

的秩均小于 $m+n$,但这是不可能的,因为 S 是可逆的.

因此,集合 $\left\{\xi^{(1)},\xi^{(2)},\cdots,\xi^{(m+n)}\right\}$ 中正好有 m 个线性无关的向量,它们是 A 的特征向量,这表明 A 必是可对角化的.同样理由,B 也是可对角化的. □

5.1.9 定理 设矩阵 $A,B\in\mathbb{C}^{n\times n}$ 是可对角化的.则 A 和 B 可交换即 $AB=BA$ 的充分必要条件是 A 和 B 同时可对角化.

证 设 A 和 B 可交换.由于 A 是可对角化的,可先对 A 和 B 执行同一相似变换将 A 对角化.

因而,不失一般性,可直接假定 A 是对角矩阵,

$$A=\mathrm{diag}(\lambda_1,\cdots,\lambda_n),$$

$\lambda_1,\cdots,\lambda_n$ 是 A 的特征值.

不失一般性,还可假定 A 的任何多重特征值是连接地出现在主对角线上的,即 A 有块对角形式:

$$A=\begin{bmatrix}\lambda_1 I & & 0\\ & \ddots & \\ 0 & & \lambda_k I\end{bmatrix}. \tag{1.8}$$

于是，从 $AB=BA$（通常的相似性不会改变这一关系）及 $B=[b_{ij}]$，得

$$\lambda_i b_{ij} = b_{ij}\lambda_j, \quad i,j=1,\cdots,n.$$

由此，当 $\lambda_i \neq \lambda_j$ 时有 $b_{ij}=0$；推出 B 是块对角矩阵：

$$B=\begin{bmatrix} B_1 & & 0 \\ & \ddots & \\ 0 & & B_k \end{bmatrix}, \tag{1.9}$$

这里 B_i 是与 $\lambda_i I$ 同阶的方阵，$i=1,\cdots,k$.

因为 B 也是可对角化的，依 **5.1.8**，每个 B_i 可对角化；设 T_i 是与 B_i 同阶的非奇异矩阵，使得

$$T_i^{-1}B_iT_i$$

是对角矩阵，$i=1,\cdots,k$.

取 T 是 $T_1,\cdots,T_k$ 的直和：

$$T=\begin{bmatrix} T_1 & & 0 \\ & \ddots & \\ 0 & & T_k \end{bmatrix}, \tag{1.10}$$

注意到

$$T_i^{-1}(\lambda_i I)T_i = \lambda_i I, \ \ i=1,\cdots,k.$$

因此 $T^{-1}AT$ 和 $T^{-1}BT$ 同为对角矩阵.

反之，如果 A 和 B 同时可对角化，那么存在相似矩阵 S，使得

$$A=SDS^{-1}, \quad B=S\Lambda S^{-1},$$

其中 D 和 Λ 是对角矩阵. 于是

$$\begin{aligned} AB &= SDS^{-1}S\Lambda S^{-1} = SD\Lambda S^{-1} \\ &= S\Lambda DS^{-1} = S\Lambda S^{-1}SDS^{-1} = BA. \end{aligned}$$ □

5.1.10 定义 任意（有限或无穷）的矩阵集合 $\mathscr{F}\subset\mathbb{C}^{n\times n}$ 也称为**矩阵族**. 如果 $\mathscr{F}$ 中的矩阵对于乘法是两两可交换的：

$$AB = BA, \quad \forall A, B \in \mathscr{F}, \tag{1.11}$$

则称 $\mathscr{F}$ 是**交换族**.

设 $A \in \mathbb{C}^{n \times n}$; $W \subset \mathbb{C}^n$ 是子空间.

如果成立

$$Aw \in W, \quad \forall w \in W,$$

则称子空间 W 为 **A-不变**的.

设族 $\mathscr{F} \subset \mathbb{C}^{n \times n}$,如果对每个 $A \in \mathscr{F}$, W 是 A-不变的,则称子空间 W 为 **$\mathscr{F}$-不变**的.

由定义即知,若 $A \in \mathbb{C}^{n \times n}$,则 $\mathbb{C}^n$ 的 1 维 A-不变子空间的每个非零元素是 A 的特征向量.

5.1.11 引理 设 $\mathscr{F} \subset \mathbb{C}^{n \times n}$ 是交换族,则存在向量 $x \in \mathbb{C}^n$,它是每个 $A \in \mathscr{F}$ 的特征向量.

证 设 $W \subset \mathbb{C}^n$ 是最小正维数的 $\mathscr{F}$-不变子空间;如此 W 是存在的,但未必唯一.因为 $\mathbb{C}^n$ 本身是 $\mathscr{F}$-不变的,也就是说,存在 n 维的 $\mathscr{F}$-不变子空间;如果还存在 $n-1$ 维的,那么可以进一步问是否存在 $n-2$ 维的;如此等等.

现在证明 W 中的每个非零向量是 $\mathscr{F}$ 中的每个矩阵的特征向量.如若不然,那么对某一矩阵 $A \in \mathscr{F}$,不是 W 中每个非零向量均为 A 的特征向量.注意到 W 是 $\mathscr{F}$-不变的,因此是 A-不变的,从而对 A 的某特征值 λ,在 W 中存在 $x \neq 0$ 使得 $Ax = \lambda x$.定义

$$W_0 \equiv \{y \in W : Ay = \lambda y\},$$

显然, $x \in W_0$, $W_0 \subset W$ 是子空间.如此,根据对 A 的假设,应有 $W_0 \neq W$, W_0 的维数小于 W 的维数.但是,对于任何 $B \in \mathscr{F}$ 有

$$Bx \in W, \quad \forall x \in W_0,$$

由此并因 $\mathscr{F}$ 是交换族,得

$$A(Bx) = (AB)x = (BA)x = B(Ax) = B(\lambda x) = \lambda(Bx),$$

推出 $Bx \in W_0$.这表明 W_0 是 $\mathscr{F}$-不变的;从而与 W 是最小正维数

$\mathscr{F}$-不变子空间矛盾. □

5.1.12 定义 $\mathscr{F} \subset \mathbb{C}^{n\times n}$ 称为**同时可对角化族**,如果存在一个非奇异矩阵 $S \in \mathbb{C}^{n\times n}$,使得对每个 $A \in \mathscr{F}$,$S^{-1}AS$ 是对角矩阵.

5.1.13 定理 设 $\mathscr{F} \subset \mathbb{C}^{n\times n}$ 是可对角化矩阵族,则 $\mathscr{F}$ 为交换族的充分必要条件是 $\mathscr{F}$ 为同时可对角化族.

此定理充分性的证明同 **5.1.9**.必要性可采用对 n 作归纳法进行论证,具体步骤从略.

5.1.14 定理 设 $A \in \mathbb{C}^{m\times n}$,$B \in \mathbb{C}^{n\times m}$,$m \le n$,则 AB 的所有特征值,按重数计,连同 $n-m$ 个 0,就是 BA 的全部特征值;也就是说,BA 和 AB 两者的特征多项式存在关系式

$$p_{BA}(\lambda) = \lambda^{n-m} p_{AB}(\lambda). \tag{1.12}$$

特别,若 $m=n$ 且 A 和 B 至少有一个是非奇异的,则 AB 和 BA 是相似的.

证 考虑关于$\mathbb{C}^{(m+n)\times(m+n)}$中块矩阵的两个恒等式

$$\begin{bmatrix} AB & 0 \\ B & 0 \end{bmatrix}\begin{bmatrix} I & A \\ 0 & I \end{bmatrix} = \begin{bmatrix} AB & ABA \\ B & BA \end{bmatrix}$$

和

$$\begin{bmatrix} I & A \\ 0 & I \end{bmatrix}\begin{bmatrix} 0 & 0 \\ B & BA \end{bmatrix} = \begin{bmatrix} AB & ABA \\ B & BA \end{bmatrix}.$$

因为

$$\begin{bmatrix} I & A \\ 0 & I \end{bmatrix} \in \mathbb{C}^{(m+n)\times(m+n)}$$

是非奇异的,推出

$$\begin{bmatrix} I & A \\ 0 & I \end{bmatrix}^{-1}\begin{bmatrix} AB & 0 \\ B & 0 \end{bmatrix}\begin{bmatrix} I & A \\ 0 & I \end{bmatrix} = \begin{bmatrix} 0 & 0 \\ B & BA \end{bmatrix},$$

所以,两个$(m+n)\times(m+n)$矩阵

$$C_1 \equiv \begin{bmatrix} AB & 0 \\ B & 0 \end{bmatrix}$$

和

$$C_2 \equiv \begin{bmatrix} 0 & 0 \\ B & BA \end{bmatrix}$$

是相似的.C_1的特征值是AB的特征值连同n个0,C_2的特征值是BA的特征值连同m个0.由于C_1和C_2的特征值相同,按重数计,定理的主要结论得证.

当$m=n$且A是非奇异时,成立

$$AB = A(BA)A^{-1},$$

表明AB和BA是相似的. □

5.2 友矩阵和 Frobenius 矩阵

5.2.1 定义 矩阵$A\in\mathbb{C}^{n\times n}$的特征方程写成

$$\begin{aligned} p_A(\lambda) &\equiv \det(\lambda I - A) \\ &= \lambda^n - p_{n-1}\lambda^{n-1} - p_{n-2}\lambda^{n-2} - \cdots - p_1\lambda - p_0 = 0. \end{aligned} \tag{2.1}$$

容易验证:矩阵

$$C \equiv \begin{bmatrix} 0 & 1 & 0 & \cdots & 0 \\ \vdots & \ddots & \ddots & \ddots & \vdots \\ \vdots & & \ddots & \ddots & 0 \\ 0 & \cdots & \cdots & 0 & 1 \\ p_0 & p_1 & \cdots & p_{n-2} & p_{n-1} \end{bmatrix} \tag{2.2}$$

的特征方程恒同 A 的特征方程(2.1).

C 称为 A 的特征多项式 $p_A(\lambda)$ 的**友矩阵**(companion matrix)或 **Frobenius 矩阵**.

友矩阵可以有不同的形式,如

$$\begin{bmatrix} p_{n-1} & p_{n-2} & \cdots & p_1 & p_0 \\ 1 & 0 & \cdots & \cdots & 0 \\ 0 & \ddots & \ddots & & \vdots \\ \vdots & \ddots & \ddots & \ddots & \vdots \\ 0 & \cdots & 0 & 1 & 0 \end{bmatrix}$$

和

$$\begin{bmatrix} 0 & \cdots & \cdots & 0 & p_0 \\ 1 & \ddots & & \vdots & p_1 \\ 0 & \ddots & \ddots & \vdots & \vdots \\ \vdots & \ddots & \ddots & 0 & p_{n-2} \\ 0 & \cdots & 0 & 1 & p_{n-1} \end{bmatrix}$$

等.

5.2.2 定理 设 $A \in \mathbf{C}^{n\times n}$ 有 n 个不同的特征值 $\lambda_1,\cdots,\lambda_n$, C 是 A 特征多项式的友矩阵.则 C 相似于对角矩阵 $\Lambda \equiv \mathrm{diag}(\lambda_1,\cdots,\lambda_n)$,而且可取相似矩阵为 $\lambda_1,\cdots,\lambda_n$ 的 Vandermonde 矩阵(见 **2.6.1**)

$$V \equiv V(\lambda_1,\cdots,\lambda_n) = \begin{bmatrix} 1 & 1 & \cdots & 1 \\ \lambda_1 & \lambda_2 & \cdots & \lambda_n \\ \lambda_1^2 & \lambda_2^2 & \cdots & \lambda_n^2 \\ \vdots & \vdots & \vdots & \vdots \\ \lambda_1^{n-1} & \lambda_2^{n-1} & \cdots & \lambda_n^{n-1} \end{bmatrix} \tag{2.3}$$

成立.

$$V^{-1}CV = \Lambda. \tag{2.4}$$

证 依 **5.2.1**, A 的特征值 $\lambda_1,\cdots,\lambda_n$ 也是 C 的特征值,因此 C 存在 n 个线性无关的特征向量.利用每个 λ_i 满足 C 的特征方程,即 $p(\lambda_i)=0$,有

$$\lambda_i^n = p_0 + p_1\lambda_i + p_2\lambda_i^2 + \cdots + p_{n-1}\lambda_i^{n-1}.$$

容易验证

$$x^{(i)} = \left(1,\lambda_i,\lambda_i^2,\cdots,\lambda_i^{n-1}\right)^{\mathrm{T}} \tag{2.5}$$

是 C 的属于特征值 λ_i 的特征向量, $x^{(i)}$ 即(2.3)中的 V 的第 i 列.于是成立 $CV=V\Lambda$,此等式等价于(2.4). □

5.2.3 定理 设 $A\in\mathbb{C}^{n\times n}$ 是非减次矩阵,有 $r\le n$ 个不同的特征值 $\lambda_1,\cdots,\lambda_r$,代数重数分别为

$$n_1,\cdots,n_r,\ \ n_1+\cdots+n_r=n,$$

则 A 相似于它的特征多项式的友矩阵 C.

证 设友矩阵 C 形如(2.2).依 **5.1.6** 和 **5.1.2**,存在非奇异矩阵 $P\in\mathbb{C}^{n\times n}$,使得

$$P^{-1}AP = J \equiv \operatorname{diag}\left(J_{n_1}(\lambda_1),\cdots,J_{n_r}(\lambda_r)\right), \tag{2.6}$$

其中 $J_{n_i}(\lambda_i)$ 由(1.3)定义.

如果 $r=n$,即代数重数

$$n_1=\cdots=n_r=1,$$

则 J 中每个 Jordan 块均为1阶矩阵,

$$J_1(\lambda_i)=[\lambda_i],\quad i=1,\cdots,n.$$

对此,**5.2.2** 已证明 J 和 C 相似,并且给出了变换矩阵 V.于是,联合(2.6),推出 A 和 C 相似.

现在考虑 J 中存在高于1阶的 Jordan 块.假设某个特征值 λ_i 的代数重数 $n_i>1$.

由于 λ_i 是形如(2.1)的特征方程 $p_A(\lambda)=0$ 的 n_i 重根,因而必须

满足

$$\frac{\mathrm{d}}{\mathrm{d}\lambda}p_A(\lambda)\bigg|_{\lambda=\lambda_i}=\binom{n}{k}\lambda_i^{n-k}-\binom{n-1}{k}p_{n-1}\lambda_i^{n-k-1}-\cdots$$

$$-\binom{k+1}{k}p_{k+1}\lambda_i-\binom{k}{k}p_k=0,$$

$$k=0,1,\cdots,n_i-1. \quad (2.7)$$

根据如此条件,只须将(2.3)的矩阵加以推广,一般地说,可取相似矩阵为

$$V=\left[V_{n_1}(\lambda_1)\quad\cdots\quad V_{n_r}(\lambda_r)\right], \quad (2.8)$$

其中$V_{n_i}(\lambda_i)$是$n\times n_i$矩阵,

$$V_{n_i}(\lambda_i)$$

$$=\begin{bmatrix} 1 & 0 & 0 & \cdots & 0 \\ \lambda_i & 1 & 0 & \cdots & 0 \\ \lambda_i^2 & \binom{2}{1}\lambda_i & 1 & \cdots & 0 \\ \lambda_i^3 & \binom{3}{1}\lambda_i^2 & \binom{3}{2}\lambda_i & \cdots & 0 \\ \vdots & \vdots & \vdots & & \vdots \\ \lambda_i^{n-2} & \binom{n-2}{1}\lambda_i^{n-3} & \binom{n-2}{2}\lambda_i^{n-4} & \cdots & \binom{n-2}{n_i-1}\lambda_i^{n-n_i-1} \\ \lambda_i^{n-1} & \binom{n-1}{1}\lambda_i^{n-2} & \binom{n-1}{2}\lambda_i^{n-3} & \cdots & \binom{n-1}{n_i-1}\lambda_i^{n-n_i} \end{bmatrix}$$

$$i=1,\cdots,r, \quad (2.9)$$

则成立

$$V^{-1}CV=J. \quad (2.10)$$

事实上,利用矩阵乘法及条件(2.7),得

$$CV_{n_i}(\lambda_i)=V_{n_i}(\lambda_i)J_{n_i}(\lambda_i)$$

$$=\begin{bmatrix} \lambda_i & 1 & 0 & \cdots & 0 \\ \lambda_i^2 & \binom{2}{1}\lambda_i & 1 & \cdots & 0 \\ \lambda_i^3 & \binom{3}{1}\lambda_i^2 & \binom{3}{2}\lambda_i & \cdots & 0 \\ \lambda_i^4 & \binom{4}{1}\lambda_i^3 & \binom{4}{2}\lambda_i^2 & \cdots & 0 \\ \vdots & \vdots & \vdots & & \vdots \\ \lambda_i^{n-1} & \binom{n-1}{1}\lambda_i^{n-2} & \binom{n-1}{2}\lambda_i^{n-3} & \cdots & \binom{n-1}{n_i-1}\lambda_i^{n-n_i} \\ \lambda_i^n & \binom{n}{1}\lambda_i^{n-1} & \binom{n}{2}\lambda_i^{n-2} & \cdots & \binom{n}{n_i-1}\lambda_i^{n-n_i+1} \end{bmatrix}$$

$$i=1,\cdots,r.$$

由此并注意到

$$CV=\left[CV_{n_1}(\lambda_1)\quad\cdots\quad CV_{n_r}(\lambda_r)\right]$$

和

$$VJ=\left[V_{n_1}(\lambda_1)J_{n_1}(\lambda_1)\quad\cdots\quad V_{n_r}(\lambda_r)J_{n_r}(\lambda_r)\right],$$

推出

$$CV=VJ,$$

这个等式等价于(2.10).至此,联合(2.6)和(2.10),即得 A 和 C 是相似的. □

更一般的是涵盖减次矩阵的结论.不难看出,当 A 是减次矩阵时,不能直接按上述定理证明中构造 V 那样来产生变换矩阵;事实上,此时 A 也的确不相似于其特征多项式的友矩阵.下面给出一般的结论.

5.2.4 Frobenius 定理 设 $A \in \mathbb{C}^{n\times n}$,则存在正整数

$$m_1,\cdots,m_s, \qquad m_1+\cdots+m_s=n,$$

使得 A 相似于矩阵

$$\hat{C}=\operatorname{diag}\left(C_{m_1},\cdots,C_{m_s}\right), \tag{2.11}$$

其中 C_{m_i} 是 $m_i\times m_i$ 矩阵,形如

$$C_{m_i}=\begin{bmatrix} 0 & 1 & 0 & \cdots & 0 \\ \vdots & \ddots & \ddots & \ddots & \vdots \\ \vdots & & \ddots & \ddots & 0 \\ 0 & \cdots & \cdots & 0 & 1 \\ p_{i,0} & p_{i,1} & \cdots & p_{i,m_i-2} & p_{i,m_i-1} \end{bmatrix},$$

$$i=1,\cdots,s, \tag{2.12}$$

每个 C_{m_i} 的特征多项式能整除其前面的 $C_{m_1},\cdots,C_{m_i-1}$ 的特征多项式.

证 设 A 的 Jordan 标准形如(1.4)和(1.5)所示,而且

(1) $J_1,\cdots,J_r$ 已按其所含的 Jordan 块的数目从多到少编号,即有 $k_1\geq\cdots\geq k_r$;

(2) 每个 J_i 内的 Jordan 块 $J_{n_i(1)}(\lambda_i),\cdots,J_{n_i(k_i)}(\lambda_i)$ 已按其阶数从高到低排列,即有 $n_i(1)\geq\cdots\geq n_i(k_i)$.

令

$$\hat{J}_1=\operatorname{diag}\left(J_{n_1(1)}(\lambda_1),J_{n_2(1)}(\lambda_2),\cdots,J_{n_r(1)}(\lambda_r)\right),$$

这就是说,$\hat{J}_1$ 是由 A 的每个特征值的最高阶的 Jordan 块组成的块对角矩阵,其阶记作 m_1, $m_1=\sum_{i=1}^{r}n_i(1)$.

类似地,令 $\hat{J}_2$ 是由 A 的每个特征值的次高阶的 Jordan 块(如果存在的话)组成的块对角矩阵,其阶记作 m_2.

如此等等,直至令 $\hat{J}_s$ 由 A 的每个特征值的第 $s=k_1$ 高阶的

Jordan 块(如果存在的话)组成的块对角矩阵,注意 $\hat{J}_s$ 不会是零阶矩阵,因为它至少包含 Jordan 块 $J_{n_1(k_1)}(\lambda_1)$,其阶记作 m_s.

显然,$m_1+\cdots+m_s=n$.这样,可以将 A 的 Jordan 标准形改写成

$$\hat{J}=\mathrm{diag}(\hat{J}_1,\cdots,\hat{J}_s), \tag{2.13}$$

这里 $\hat{J}_1,\cdots,\hat{J}_s$ 均必为非减次矩阵.

现在对每个 $\hat{J}_i$ 可以应用 **5.2.3** 设形如(2.12)的 C_{m_i} 是 $\hat{J}_i$ 的特征多项式的友矩阵,则存在 $m_i\times m_i$ 非奇异矩阵 V_i,使得

$$V_i^{-1}C_{m_i}V_i=\hat{J}_i,\quad i=1,\cdots,s. \tag{2.14}$$

于是成立

$$\hat{V}^{-1}\hat{C}\hat{V}=\hat{J}, \tag{2.15}$$

其中 $\hat{C}$ 形如(2.11),$\hat{J}$ 形如(2.13),而

$$\hat{V}=\mathrm{diag}(V_1,\cdots,V_s), \tag{2.16}$$

(2.15)表明 $\hat{J}$ 和 $\hat{C}$ 相似.然而,显然 A 和 $\hat{J}$ 是相似的,因此 A 相似于 $\hat{C}$.

从 $\hat{J}_1,\cdots,\hat{J}_s$ 的构造得知:每个 C_{m_i} 的特征多项式能整除其前面的 $C_{m_1},\cdots,C_{m_i-1}$ 的特征多项式. □

这一定理的另一证明见 **9.6.16**.

5.3 Schur 标准形

5.3.1 引言 如果能把一个矩阵约化成 Jordan 标准形,那么便可得到特征值和特征向量的全部信息;但是在实际约化方面,除了如自伴矩阵等一些类型外,缺乏普遍有效的变换矩阵.

自伴矩阵总是可以通过酉相似变换(即利用酉矩阵作相似变换)将其约化成对角形,即其 Jordan 标准形.

酉变换具有很引人的一些性质.因此,自然会问:酉相似变换能将一般矩阵约化到什么程度呢?下面将给出的答案是:如此变换可将任何矩阵约化为三角矩阵.

三角矩阵保留了对角矩阵的一个良好性质,即其特征值分布在主对角线,而且求对应的特征向量也比较简单.因而约化成三角标准形在理论上和应用上都很重要. Jordan 标准形实际上也是三角标准形的特殊情形.

5.3.2 引理 设 $x=(x_1,\cdots,x_n)^{\mathrm{T}}\in\mathbb{C}^n$, $x\neq 0$,令

$$H=I-2ww^*, \tag{3.1}$$

其中

$$w=\frac{x-pe_1}{\|x-pe_1\|},\quad p\in\mathbb{C},\quad e_1=(1,0,\cdots,0)^{\mathrm{T}}, \tag{3.2}$$

则 H 是自伴的酉矩阵,

$$H^*=H=H^{-1}. \tag{3.3}$$

而且,取

$$p=\begin{cases}\|x\|_2, & x_1=0,\\ -\mathrm{e}^{\mathrm{i}\arg x_1}\|x\|_2, & x_1\neq 0\end{cases} \tag{3.4}$$

时,成立

$$Hx=pe_1. \tag{3.5}$$

证 显然,$H^*=H$;又 $w^*w=1$,

$$\begin{aligned}H^*H&=(I-2ww^*)(I-2ww^*)\\&=I-4ww^*+4ww^*ww^*\\&=I-4ww^*+4ww^*=I,\end{aligned}$$

(3.3)得证.

因为

$$
\begin{aligned}
Hw &= x - 2ww^*x \\
&= x - \frac{2}{\|x - pe_1\|_2^2}(x - pe_1)(x^* - p^*e_1^*)x \\
&= x - \frac{2\left(\|x\|_2^2 - p^*x_1 \right)}{\|x - pe_1\|_2^2}(x - pe_1),
\end{aligned}
$$

所以,欲使(3.5)成立,只须

$$
\|x - pe_1\|_2^2 = 2\left(\|x\|_2^2 - p^*x_1 \right).
$$

由此,将其中的2－范数具体写成分量形式,即可得出条件(3.4).这样,反过来说,在条件(3.4)下自然成立(3.5). □

5.3.3 Schur 上三角标准形定理 设 $A \in \mathbb{C}^{n\times n}$,则存在酉矩阵 $U \in \mathbb{C}^{n\times n}$,使得

$$
U^*AU = \Delta \tag{3.6}
$$

其中 Δ 是上三角矩阵,其对角元素(即 A 的特征值)可按指定的顺序排列.

特别,若 $A \in \mathbb{R}^{n\times n}$ 且 A 的特征值是实的,则 U 可以选取为实的和正交的.

Δ 称为 A 的 **Schur 上三角标准形**.

(3.6)的等价形式

$$
A = U\Delta U^*,
$$

称为 A 的 **Schur 分解**.

证 利用数学归纳法.

$n = 1$ 时结论显然成立.

现在考虑 $n \geq 2$.假设结论对所有 $n-1$ 阶矩阵成立.而且欲将 A 的一个特征值 λ 排列在 Δ 的左上角.用 x 表示相应于 λ 的特征向量,

$$
Ax = \lambda x.
$$

依 **5.3.2**,存在自伴的酉矩阵 H 和常数 $p \neq 0$,使

$$
Hx = pe_1.
$$

于是

$$HAHe_1 = HAH\left(\frac{1}{p}Hx\right) = \frac{1}{p}HAx = \frac{\lambda}{p}Hx = \lambda e_1,$$

这表明 HAH 形如

$$HAH = \begin{bmatrix} \lambda & u \\ 0 & A_1 \end{bmatrix},$$

其中 A_1 是 $n-1$ 阶矩阵, A_1 及 $n-1$ 维行向量 u 的元素对这里的推证并不重要.

注意到 $H = H^{-1}$, A 和 HAH 相似,故 A_1 的特征值包括了 A 的除 λ 外的全部特征值.

依归纳法假设,存在 $n-1$ 阶酉矩阵 U_1,使得

$$U_1^* A_1 U_1 = \Delta_1,$$

其中 Δ_1 是 $n-1$ 阶上三角矩阵,其对角元素已按指定顺序排列好.最后令

$$U = H\text{diag}(1, U_1),$$

显然, U 是 n 阶酉矩阵,而且

$$U^*AU = \begin{bmatrix} 1 & 0 \\ 0 & U_1^* \end{bmatrix}\begin{bmatrix} \lambda & u \\ 0 & A_1 \end{bmatrix}\begin{bmatrix} 1 & 0 \\ 0 & U_1 \end{bmatrix} = \begin{bmatrix} \lambda & uU_1 \\ 0 & \Delta_1 \end{bmatrix} \equiv \Delta.$$

若 $A \in \mathbb{R}^{n\times n}$ 且 A 的特征值是实的,则对应的特征向量可选为实的,因而上述所有步骤可以在实运算中实现. □

注: Schur 分解也可以取下三角矩阵形式的标准形.

5.3.4 定理 设 $\mathscr{F} \subset \mathbb{C}^{n\times n}$ 是交换族,则存在酉矩阵 $U \in \mathbb{C}^{n\times n}$,使得对于每个 $A \in \mathscr{F}$, U^*AU 是三角矩阵.

证 回到 **5.3.3** 的证明.

利用 **5.1.11** 证明中选取特征向量(从而酉矩阵)的每一步,对于每个 $A \in \mathscr{F}$,可以选取相同的特征向量(从而相同的酉矩阵).而且,酉等价不改变交换性;两个形如

$$\begin{bmatrix} A_{11} & A_{12} \\ 0 & A_{22} \end{bmatrix} \quad 和 \quad \begin{bmatrix} B_{11} & B_{12} \\ 0 & B_{22} \end{bmatrix}$$

的相匹配的分块矩阵,如果是可交换的,那么 A_{22} 和 B_{22} 也是可交换的.

因此,在 **5.3.3** 的应用归纳法的约化过程中,交换族的性质能得以遗传.这样,对于 $\mathscr{F}$ 的所有成员来说,**5.3.3** 中关于 U 的各个组成部分都可以按相同的方式选取. □

5.3.5 定理 设 $A\in\mathbb{R}^{n\times n}$,则存在实正交矩阵 $Q\in\mathbb{R}^{n\times n}$,使得

$$Q^{\mathrm{T}}AQ=\begin{bmatrix} A_1 & * & \cdots & * \\ 0 & A_2 & \ddots & \vdots \\ \vdots & \ddots & \ddots & * \\ 0 & \cdots & 0 & A_k \end{bmatrix}\in\mathbb{R}^{n\times n}, \tag{3.7}$$

其中 $1\le k\le n$;每个 A_i 是 1×1 实矩阵,或具有一对非实的复共轭特征值的 2×2 实矩阵.而且,对角块 $A_1,\cdots,A_k$ 可以按任何指定的顺序排列.

这一定理是 **5.3.3** 关于实情形的说法.

5.3.6 定理 设 $\mathscr{F}\subset\mathbb{R}^{n\times n}$ 是交换族,则存在实正交矩阵 $Q\in\mathbb{R}^{n\times n}$,使得对于每个 $A\in\mathscr{F}$,$Q^{\mathrm{T}}AQ$ 形如(3.7).

这一定理是 **5.3.4** 关于实情形的说法.

5.3.7 定理 设 $A=[a_{ij}]\in\mathbb{C}^{n\times n}$,则对每个正数 $\varepsilon>0$,存在具有 n 个不同的特征值(从而可对角化)的矩阵

$$A(\varepsilon)=[a_{ij}(\varepsilon)]\in\mathbb{C}^{n\times n},$$

使得

$$\sum_{i,j=1}^{n}\left|a_{ij}-a_{ij}(\varepsilon)\right|^2<\varepsilon. \tag{3.8}$$

证 设$U\in\mathbb{C}^{n\times n}$是酉矩阵,使得

$$U^*AU=T\equiv[t_{ij}]$$

是上三角矩阵.选取$D=\mathrm{diag}(\alpha_1,\cdots,\alpha_n)$,

$$|\alpha_i|<\left(\frac{\varepsilon}{n}\right)^{1/2},\qquad i=1,\cdots,n,$$

而且使$t_{11}+\alpha_1,\cdots,t_{nn}+\alpha_n$这$n$个数互不相同.因此$T+D$以及与其相似的$A+UDU^*$有$n$个不同特征值:

$$t_{11}+\alpha_1,\cdots,t_{nn}+\alpha_n.$$

取$A(\varepsilon)=A+UDU^*$,于是

$$A-A(\varepsilon)=-UDU^*,$$

从而

$$\sum_{i,j=1}^{n}\left|a_{ij}-a_{ij}(\varepsilon)\right|^2=\sum_{i=1}^{n}|\alpha_i|^2<n\left(\frac{\varepsilon}{n}\right)=\varepsilon.$$ □

5.3.8 定理 设$A\in\mathbb{C}^{n\times n}$,则对每个正数$\varepsilon>0$,存在非奇异矩阵$S_\varepsilon\in\mathbb{C}^{n\times n}$,使得

$$S_\varepsilon^{-1}AS_\varepsilon=T_\varepsilon\equiv\left[t_{ij}(\varepsilon)\right] \tag{3.9}$$

是上三角矩阵,而且

$$\left|t_{ij}(\varepsilon)\right|<\varepsilon,\quad 1\le i<j\le n.$$

证 依5.3.3,存在酉矩阵$U\in\mathbb{C}^{n\times n}$和上三角矩阵

$$T=[t_{ij}]\in\mathbb{C}^{n\times n},$$

满足

$$U^*AU=T.$$

定义

$$D_\alpha = \mathrm{diag}\left(1, \alpha, \alpha^2, \cdots, \alpha^{n-1}\right),\ \alpha \neq 0.$$

令

$$t = \max_{i<j} \left| t_{ij} \right|,$$

而且不妨假设 $\varepsilon < 1$.

如果 $t \le 1$,取 $S_\varepsilon = UD_\varepsilon$,则

$$S_\varepsilon^{-1} A S_\varepsilon = D_\varepsilon^{-1} U^{-1} A U D_\varepsilon = D_\varepsilon^{-1} U^* A U D_\varepsilon = D_\varepsilon^{-1} T D_\varepsilon \equiv T_\varepsilon,$$

从而

$$\left| t_{ij}(\varepsilon) \right| = \left| t_{ij} \varepsilon^{-i+1} \varepsilon^{j-1} \right| = \left| t_{ij} \right| \varepsilon^{j-i} \le \varepsilon^{j-i} \le \varepsilon,\ \forall i < j.$$

如果 $t > 1$,取 $S_\varepsilon = UD_{1/t} D_\varepsilon$,则类似地,

$$S_\varepsilon^{-1} A S_\varepsilon = D_\varepsilon^{-1} D_{1/t}^{-1} T D_{1/t} D_\varepsilon \equiv T_\varepsilon,$$

从而

$$\left| t_{ij}(\varepsilon) \right| = \left| t_{ij} \right| \varepsilon^{j-i} \left(\frac{1}{t} \right)^{j-i}$$

$$\le t \left(\frac{1}{t} \right)^{j-i} \varepsilon^{j-i} \le \varepsilon,\ \forall i < j. \qquad \square$$

5.3.7 和 **5.3.8** 从两种意义上表明每个矩阵是**近乎可对角化的**(almost diagonalizable).**5.3.7** 是说对于任何给定矩阵,存在接近于它到任意指定程度的可对角化矩阵;**5.3.7** 则是说任何给定矩阵必相似于非对角元素小于任意指定程度的上三角矩阵.

5.3.9 定理 设 $A \in \mathbb{C}^{n\times n}$ 有 n_i 重特征值 $\lambda_i, i = 1, \cdots, k$, $\lambda_1, \cdots, \lambda_k$ 互不相同,则 A 相似于如下形式的矩阵:

$$\begin{bmatrix} T_1 & & & 0 \\ & T_2 & & \\ & & \ddots & \\ 0 & & & T_k \end{bmatrix}, \tag{3.10}$$

其中$T_i \in \mathbb{C}^{n_i \times n_i}$是对角元素全为$\lambda_i$的上三角矩阵,$i = 1, \cdots, k$.

特别,若$A \in \mathbb{R}^{n \times n}$而且所有特征值是实的,则相似矩阵可取为实的.

证 首先依 **5.3.3**, A酉相似于上三角矩阵$T = [t_{rs}] \in \mathbb{C}^{n \times n}$,并假设$T$的对角线已然所有$\lambda_1$排在最前,其次是所有$\lambda_2$,等等.

现在对T执行一系列简单(非酉)相似变换,以产生所希望的上对角线中的零元素,而不改变T的对角元素和上三角结构.令$E_{rs} \in \mathbb{C}^{n \times n}$是$(r,s)$位置为1,其余位置为0的矩阵.注意到对于$r \neq s$和任意的$\alpha \in \mathbb{C}$, $I + \alpha E_{rs}$是非奇异的,而且

$$(I + \alpha E_{rs})^{-1} = I - \alpha E_{rs}.$$

因此,对于$r < s$,如果用$I + \alpha E_{rs}$作为相似矩阵,那么

$$(I + \alpha E_{rs})^{-1} T (I + \alpha E_{rs}) = (I - \alpha E_{rs}) T (I + \alpha E_{rs})$$

仅改变T的如下位置的元素:

$$\begin{array}{ccccccc}
(1,s) & & & & & & \\
\downarrow & & & & & & \\
(2,s) & & & & & & \\
\downarrow & & & & & & \\
\vdots & & & & & & \\
\downarrow & & & & & & \\
(r,s) & \rightarrow & (r,s+1) & \rightarrow & \cdots & \rightarrow & (r,n).
\end{array}$$

而且将t_{rs}替换成

$$t_{rs} + \alpha(t_{rr} - t_{ss}),$$

因此,当$t_{rr} \neq t_{ss}$时,选取

$$\alpha = \frac{-t_{rs}}{t_{rr} - t_{ss}}$$

可使(r,s)位置的元素为0,此外有关结构没有改变.于是,考虑T中的如下位置序列:

$$
\begin{aligned}
&(n-1,n);\\
&(n-2,n-1),(n-2,n);\\
&(n-3,n-2),(n-3,n-1),(n-3,n);\\
&\cdots\cdots\\
&(1,2),(1,3),\cdots,(1,n).
\end{aligned}
$$

只要$t_{rr}\neq t_{ss}$,便可以按所描述的方式选取相似矩阵,依次使以上位置中的有关元素为0,注意先前产生的0元素不会受到影响.所得矩阵必相似于A而且形如(3.10). □

此定理是**5.3.3**的延伸,是趋近Jordan标准形的重要一步.

5.3.10 定理 设$A,B\in\mathbb{C}^{n\times n}$,$A$和$B$可交换,而且特征值分别为$\lambda_1,\cdots,\lambda_n$和$\mu_1,\cdots,\mu_n$,则存在指标$1,\cdots,n$的一个置换

$$i_1,\cdots,i_n$$

使得$A+B$的特征值是

$$\lambda_1+\mu_{i_1},\cdots,\lambda_n+\mu_{i_n}.$$

笼统地说,若A和B可交换,则

$$\lambda(A+B)\subset\lambda(A)+\lambda(B). \tag{3.11}$$

证 依**5.3.4**,A和B同时可上三角化,即存在酉矩阵$U\in\mathbb{C}^{n\times n}$,使得

$$U^*AU=T,\ U^*BU=R,$$

T和R同为上三角矩阵,对角元素分别为$\lambda_1,\cdots,\lambda_n$和$\mu_1,\cdots,\mu_n$.显然

$$U^*(A+B)U=T+R,$$

因此,$A+B$和$T+R$相似,必有相同的特征值.而$T+R$是上三角矩阵,特征值为$\lambda_1+\mu_{i_1},\cdots,\lambda_n+\mu_{i_n}$. □

5.3.11 推论 设 $A,B\in\mathbb{C}^{n\times n}$, A 和 B 可交换,而且特征值分别为 $\lambda_1,\cdots,\lambda_n$ 和 $\mu_1,\cdots,\mu_n$.如果

$$\lambda_i\neq-\mu_j,\ \ i,j=1,\cdots,n,$$

那么 $A+B$ 是非奇异的. □

5.4 奇异值分解

5.4.1 定义 设 $A\in\mathbb{C}^{m\times n}$,由于 $n\times n$ 矩阵 A^*A 是半正定的,其特征值的非负平方根称为 A 的**奇异值**,记作 $\sigma_1\geq\cdots\geq\sigma_n\geq0$.并用 $\sigma(A)$表示 A 的奇异值的全体:

$$\sigma(A)\equiv\{\sigma\geq0:A^*Ax=\sigma^2x,x\in\mathbb{R}^n,x\neq0\}. \tag{4.1}$$

奇异值和奇异值分解起因于19世纪微分几何学家和代数学家提出的如下问题:设

$$A\equiv[a_{ij}],B\equiv[b_{ij}]\in\mathbb{R}^{n\times n},$$

而且

$$x=(x_1,\cdots,x_n)^{\mathrm{T}},y=(y_1,\cdots,y_n)^{\mathrm{T}}\in\mathbb{R}^n.$$

怎样判断两个实双线性型

$$\varphi(x,y)=\sum_{i,j=1}^{n}a_{ij}x_iy_j \tag{4.2}$$

和

$$\psi(x,y)=\sum_{i,j=1}^{n}b_{ij}x_iy_j \tag{4.3}$$

在独立的实正交替换下是否等价,也就是说,是否存在实正交矩阵 $Q_1,Q_2\in\mathbb{R}^{n\times n}$,使得

$$\varphi(x,y)=\psi(Q_1x,Q_2y),\ \ \forall x,y\in\mathbb{R}^n.$$

解决此问题的一种途径是寻求双线性型在正交替换下的标准形,而且发现了每个 $A\in\mathbb{R}^{n\times n}$ 必存在正交矩阵 $Q_1,Q_2\in\mathbb{R}^{n\times n}$,使得

$$Q_1^{\mathrm{T}}AQ_2=\operatorname{diag}(\sigma_1,\cdots,\sigma_n) \tag{4.4}$$

是非负的对角矩阵,其中

$$\sigma_1^2\geq\cdots\geq\sigma_n^2$$

是 $A^{\mathrm{T}}A$(也是 AA^{T})的特征值.用现在的术语来说,(4.4)就是实方阵的奇异值分解,而

$$\sigma_1\geq\cdots\geq\sigma_n\geq 0$$

是 A 的奇异值.

现在奇异值和奇异值分解在一些应用领域和理论领域中起非常重要的作用:诸如高质量统计计算,基于用低秩矩阵逼近给定矩阵的数据压缩方案,酉不变范数理论等.

5.4.2 定理 设 $A\in\mathbb{C}^{m\times n}$.则 A 和 A^* 有相同的非零奇异值:

$$\sigma_1\geq\cdots\geq\sigma_r>0,\ \ r\leq q=\min\{m,n\}.$$

证 设 $\sigma>0$ 是 A 的奇异值,亦即 σ^2 是 A^*A 的特征值:

$$A^*Ax=\sigma^2x,$$

其中 $x\in\mathbb{R}^n$,$x\neq 0$,是 A^*A 相应于 σ^2 的特征向量.因而,

$$(Ax,Ax)=(A^*Ax,x)=\sigma^2(x,x).$$

由此,因 $\sigma^2>0$,$(x,x)\neq 0$,推出 $Ax\neq 0$.于是

$$(AA^*)(Ax)=A(A^*Ax)=\sigma^2(Ax),$$

这表明 σ^2 是

$$(A^*)^*(A^*)=AA^*$$

的特征值,而 Ax 是相应的特征向量.因此 σ 是 A^* 的奇异值. □

A 和 A^* 的非零奇异值的个数

$$r = \mathrm{rank}(A^*A) = \mathrm{rank}(AA^*). \tag{4.5}$$

上述定理表明,当 $n \neq m$ 时, A 和 A^* 的奇异值仅有零奇异值的重数不同. A 的零奇异值的重数(即 A^*A 的特征值的几何重数)为 $n-r$, 而 A^* 的零奇异值的重数(即 AA^* 的特征值的几何重数)则为 $m-r$. 自然,对于方阵($n=m$), A 和 A^* 的奇异值是重合的,而且有相同的重数.

5.4.3 定理 方阵的奇异值在酉变换下不变.具体地说,对于任意的 $A \in \mathbb{C}^{n\times n}$ 和任意的酉矩阵 $U \in \mathbb{C}^{n\times n}$,成立

$$\sigma(A) = \sigma(UA) = \sigma(AU). \tag{4.6}$$

证 设 $\sigma \in \sigma(A)$,依 **5.4.1**,存在 $x \in \mathbb{R}^n$, $x \neq 0$,

$$A^*Ax = \sigma^2 x.$$

由此,有

$$(UA)^*(UA)x = A^*U^*UAx = A^*Ax = \sigma^2 x,$$

推出 $\sigma \in \sigma(UA)$;类似地,

$$(AU)^*(AU)(U^*x) = U^*A^*Ax = \sigma^2(U^*x),$$

注意到 $U^*x \neq 0$,推出 $\sigma \in \sigma(AU)$. □

5.4.4 定理 设 $A \in \mathbb{C}^{n\times n}$ 是正规矩阵,则 A^* 的特征值全体

$$\lambda(A^*) = \{\bar{\lambda} : \lambda \in \lambda(A)\}. \tag{4.7}$$

A 和 A^* 的奇异值全体

$$\sigma(A) = \sigma(A^*) = \{|\lambda| : \lambda \in \lambda(A)\}. \tag{4.8}$$

证 因为 $A^*A = AA^*$,即 A 和 A^* 可交换,依 **5.3.4**,存在酉矩阵 $U \in \mathbb{C}^{n\times n}$,使得

$$U^*AU = T, \quad U^*A^*U = R,$$

其中 T 和 R 同为上三角矩阵.因此,有

$$R^* = (U^*A^*U)^* = U^*AU = T.$$

注意到 R^* 是下三角矩阵,T 是上三角矩阵,两者相等必同为对角矩阵:

$$R^* = T = \operatorname{diag}(\lambda_1, \cdots, \lambda_n),$$

这里 $\lambda_1, \cdots, \lambda_n$ 是 A 的特征值,即 $\lambda(A) = \{\lambda_1, \cdots, \lambda_n\}$.于是

$$U^* A^* U = R = \operatorname{diag}(\overline{\lambda_1}, \cdots, \overline{\lambda_n}),$$

这表明(4.7)成立;而且

$$\begin{aligned} U^* A^* A U &= (U^* A^* U)(U^* A U) \\ &= RT = \operatorname{diag}(|\lambda_1|^2, \cdots, |\lambda_n|^2), \end{aligned}$$

推出(4.8). □

容易证明下列两条性质(留作练习):

(1) 设 $A \in \mathbb{C}^{n\times n}$,$\sigma(A) = \{\sigma_1, \cdots, \sigma_n\}$,则

$$|\det A| = \sigma_1 \cdots \sigma_n. \tag{4.9}$$

(2) 方阵为酉矩阵的充分必要条件是其奇异值全为1.

5.4.5 奇异值分解定理 设 $A \in \mathbb{C}^{m\times n}$ 的秩为 r,则存在酉矩阵 $U \in \mathbb{C}^{m\times m}$ 和酉矩阵 $V \in \mathbb{C}^{n\times n}$,使得

$$U^* A V = \Sigma \equiv \begin{bmatrix} \Sigma_r & 0_3 \\ 0_2 & 0_1 \end{bmatrix}, \tag{4.10}$$

其中

$$\Sigma_r = \operatorname{diag}(\sigma_1, \cdots, \sigma_r), \tag{4.11}$$

$\sigma_1 \geq \cdots \geq \sigma_r > 0$ 是 A 的全部非零奇异值;而 $0_1, 0_2, 0_3$ 分别是 $(m-r)\times(n-r)$,$(m-r)\times r$,$r\times(n-r)$ 的零矩阵.

证 因为 $A^* A$ 是 $n \times n$ 半正定矩阵,依 **5.3.3**,它有如下 Schur 分解

$$V^* (A^* A) V = \operatorname{diag}(\sigma_1^2, \cdots, \sigma_n^2), \tag{4.12}$$

其中 $V \in \mathbb{C}^{n\times n}$ 是酉矩阵,将其表示成

$$V=\begin{bmatrix}V_1 & V_2\end{bmatrix},$$

其中

$$V_1=\begin{bmatrix}v_1 & \cdots & v_r\end{bmatrix},\quad V_2=\begin{bmatrix}v_{r+1} & \cdots & v_n\end{bmatrix},$$

v_i 是 V 的第 i 列,$i=1,\cdots,n$.从

$$\operatorname{rank}\left(A^*A\right)=\operatorname{rank}A=r,$$

得

$$\sigma_1\geq\cdots\geq\sigma_r>0,\ \sigma_{r+1}=\cdots=\sigma_n=0.$$

令 Σ_r 形如(4.11),从(4.12)有

$$V_1^*\left(A^*A\right)V_1=\Sigma_r^2,\quad V_2^*\left(A^*A\right)V_2=0,$$

这表明 AV_1 的列相互正交;AV_2 的列是零向量,即 $AV_2=0$.因此,若取 $m\times r$ 矩阵

$$U_1\equiv AV_1\Sigma_r^{-1},$$

则 $U_1^*U_1=I$.

依 **1.7.9**,U_1 的列这一组正交向量可以扩展成 $\mathbb{C}^m$ 中的一组正交基,故存在 $m\times(m-r)$ 矩阵 U_2,使得

$$U=\begin{bmatrix}U_1 & U_2\end{bmatrix}\in\mathbb{C}^{m\times m}$$

是酉矩阵.于是,$U_2^*AV_1=U_2^*U_1\Sigma_r=0$,

$$U^*AV=\begin{bmatrix}U_1^*AV_1 & U_1^*AV_2\\ U_2^*AV_1 & U_2^*AV_2\end{bmatrix}=\begin{bmatrix}\Sigma_r & 0\\ 0 & 0\end{bmatrix}.$$ □

(4.10)或等价地 $A=U\Sigma V^*$ 称为 A 的**奇异值分解**.

V 的第 i 列 $(1\leq i\leq n)$

$$v_i=Ve_i$$

称为 A 的属于奇异值 σ_i 的**单位右奇异向量**.

U 的第 i 列 $(1\leq i\leq m)$

$$u_i=Ue_i$$

称为 A 的属于奇异值 σ_i 的**单位左奇异向量**.

一般,(4.10)中的 Σ_r 是唯一的,但每个奇异值 σ_i 对应的单位奇异向量是不唯一的.

从(4.10)得

$$Av_i=\sigma_i u_i,\quad A^*u_i=\sigma_i v_i,\quad i=1,\cdots,r. \tag{4.13}$$

奇异值分解 $A=U\Sigma V^*$ 是自伴或正规矩阵谱分解 $A=U\Lambda U^*$ 的自然推广,因此自伴或正规矩阵的常见性质可能引出关于一般矩阵的有用结果.

5.4.6 定理 设 $A\in\mathbb{C}^{m\times n}$, $\sigma_1\geq\cdots\geq\sigma_n$ 是 A 的有序奇异值,则对于 $k=1,\cdots,\min\{m,n\}$,成立

$$\sigma_k=\min_{\dim S=k-1}\ \max_{x\perp S,\|x\|_2=1}\|Ax\|_2 \tag{4.14}$$

$$=\max_{\dim S=n-k}\ \min_{x\perp S,\|x\|_2=1}\|Ax\|_2 \tag{4.15}$$

$$=\min_{\dim S=n-k+1}\ \max_{x\in S,\|x\|_2=1}\|Ax\|_2 \tag{4.16}$$

$$=\max_{\dim S=k}\ \min_{x\in S,\|x\|_2=1}\|Ax\|_2, \tag{4.17}$$

其中外优化是对指定维数的所有子空间 $S\in\mathbb{C}^n$ 而言的.

证 将3.5.3的证明应用于 A^*A 的递减有序特征值

$$\lambda_1\geq\cdots\geq\lambda_n\geq 0,$$

有

$$\lambda_k=\min_{\dim S=n-k+1}\ \max_{x\in S,x\neq 0}\frac{(x,A^*Ax)}{(x,x)}$$

$$=\min_{\dim S=n-k+1}\ \max_{x\in S,\|x\|_2=1}\|Ax\|_2^2.$$

由此,并注意到 $\lambda_k=\sigma_k^2$,即得(4.16)和(4.14).

类似地,可以证明(4.17)和(4.15). □

5.4.7 推论 设 $A\in\mathbb{C}^{m\times n}$，$\sigma_1\ge\sigma_2\ge\cdots$ 是 A 的奇异值，用 A_r 表示从 A 删去 r 行(与/或列)而得到的子矩阵，并用 $\hat{\sigma}_1\ge\hat{\sigma}_2\ge\cdots$ 表示 A_r 的奇异值，则

$$\sigma_k\ge\hat{\sigma}_k\ge\sigma_{k+r},\quad k=1,\cdots,\min\{m,n\},\tag{4.18}$$

这里，对于 $X\in\mathbb{C}^{p\times q}$，规定当 $j>\min\{p,q\}$ 时 X 的奇异值 $\sigma_j=0$.

证 只须考虑 $r=1$ 的情形.事实上，从 A 删去1行或1列，如果成立 $\sigma_k\ge\hat{\sigma}_k\ge\sigma_{k+1}$，那么反复应用如此不等式便可得出一般结论(4.18).

现在，假设 A_1 是从 A 删去第 s 列形成的.对于向量 $x\in\mathbb{C}^n$，用 $\hat{x}\in\mathbb{C}^{n-1}$ 表示 x 删去第 s 个分量后的向量.于是，利用(4.14)有

$$\begin{aligned}\sigma_k&=\min_{\dim S=k-1}\max_{x\perp S,\|x\|_2=1}\|Ax\|_2\\&\ge\min_{\dim S=k-1}\max_{x\perp S,x\perp e_s,\|x\|_2=1}\|Ax\|_2\\&=\min_{\dim S=k-1}\max_{\hat{x}\perp S,\|\hat{x}\|_2=1}\|A_1\hat{x}\|_2=\hat{\sigma}_k.\end{aligned}$$

再利用(4.15)有

$$\begin{aligned}\sigma_{k+1}&=\max_{\dim S=n-k-1}\min_{x\perp S,\|x\|_2=1}\|Ax\|_2\\&\le\max_{\dim S=n-k-1}\min_{x\perp S,x\perp e_s,\|x\|_2=1}\|Ax\|_2\\&=\max_{\dim S=n-k-1}\min_{\hat{x}\perp S,\|\hat{x}\|_2=1}\|A_1\hat{x}\|_2=\hat{\sigma}_k.\end{aligned}$$

若删去 A 的第 s 行，可将同样的推证应用于 A^*，而 A^* 与 A 有同样的奇异值. □

5.4.8 推论 设 $A\in\mathbb{C}^{n\times n}$，$\sigma_1\ge\cdots\ge\sigma_n$ 是 A 的有序奇异值；又设 $H\equiv\frac{1}{2}\left(A+A^*\right)$ 是 A 的自伴部分，$\lambda_1\ge\cdots\ge\lambda_n$ 是 H 的递减有序特征值，则

$$\sigma_k \ge \lambda_k, \quad k=1,\cdots,n. \tag{4.19}$$

更一般地,对任何酉矩阵$U,V \in \mathbb{C}^{n\times n}$,用$\hat{\lambda}_1 \ge \cdots \ge \hat{\lambda}_n$表示$UAV$的自伴部分的特征值,成立

$$\sigma_k \ge \hat{\lambda}_k, \quad k=1,\cdots,n. \tag{4.20}$$

证 对任何$x \in \mathbb{C}^n$,$\|x\|_2 = 1$,有

$$x^* Hx = \frac{1}{2}\left(x^* Ax + x^* A^* x\right) = \mathrm{Re}\left(x^* Ax\right)$$
$$\le \left|x^* Ax\right| \le \|x\|_2 \|Ax\|_2 = \|Ax\|_2 .$$

因此,利用(4.14),得

$$\lambda_k = \min_{\dim S = k-1} \max_{x \perp S, \|x\|_2 = 1} x^* Hx$$
$$\le \min_{\dim S = k-1} \max_{x \perp S, \|x\|_2 = 1} \|Ax\|_2 = \sigma_k .$$

更一般的结论(4.20),则可从UAV和A显然有相同的奇异值而推出. □

5.4.9 定义 $P \in \mathbb{C}^{m\times n}$称为**秩$r$部分等距矩阵**,如果$P$的奇异值

$$\sigma_1 = \cdots = \sigma_r = 1, \ \sigma_{r+1} = \cdots = \sigma_q = 0, \tag{4.21}$$

其中$q = \min\{m,n\}$.

两个部分等距矩阵$P,Q \in \mathbb{C}^{m\times n}$称为**正交的**,如果$P^*Q = 0$且$PQ^* = 0$.

5.4.10 定理 设$A \in \mathbb{C}^{m\times n}$的奇异值分解为$A = U\Sigma V^*$,其中

$$U = \begin{bmatrix} u_1 & \cdots & u_m \end{bmatrix} \in \mathbb{C}^{m\times m}$$

和

$$V = \begin{bmatrix} v_1 & \cdots & v_n \end{bmatrix} \in \mathbb{C}^{n\times n}$$

是酉矩阵,而$\Sigma = [\sigma_{ij}] \in \mathbb{C}^{m\times n}$具有

$$\sigma_1 \equiv \sigma_{11} \geq \cdots \geq \sigma_q \equiv \sigma_{qq}, \quad q = \min\{m, n\},$$

其余$\sigma_{ij} = 0$,则

(1) A是相互正交的秩1部分等距矩阵

$$P_i = u_i v_i^*, \quad i = 1, \cdots, q$$

的非负线性组合

$$A = \sigma_1 P_1 + \cdots + \sigma_q P_q. \tag{4.22}$$

(2) A可以表示为下列部分等距矩阵$K_1, \cdots, K_q$的非负线性组合:

$$A = \mu_1 K_1 + \cdots + \mu_q K_q, \tag{4.23}$$

这里

$$\begin{aligned} &K_i = UEV^*; \\ &E_i \equiv [e_1 \quad \cdots \quad e_i \quad 0 \quad \cdots \quad 0] \in \mathbb{R}^{m \times n}; \\ &\operatorname{rank} K_i = i, \qquad\qquad i = 1, \cdots, q. \end{aligned} \tag{4.24}$$

并且

$$\mu_i = \sigma_i - \sigma_{i+1}, \quad i = 1, \cdots, q-1; \quad \mu_q = \sigma_q, \tag{4.25}$$

因而

$$\mu_1 + \cdots + \mu_q = \sigma_i, \ i = 1, \cdots, q. \tag{4.26}$$

这一定理所含结论的证明是直截了当的,留作练习.

5.4.11 极分解定理 设$A \in \mathbb{C}^{m \times n}$.

(1) 若$n \geq m$,则

$$A = PY,$$

其中$P \in \mathbb{C}^{m \times m}$是半正定的,$P^2 = AA^*$;而$Y \in \mathbb{C}^{m \times n}$是行正交的,即$YY^* = I \in \mathbb{C}^{m \times m}$.

(2) 若$m \geq n$,则

$$A = XQ,$$

其中$Q \in \mathbb{C}^{n \times n}$是半正定的,$Q^2 = A^*A$;而$X \in \mathbb{C}^{m \times n}$是列正交的,

即 $X^*X = I \in \mathbb{C}^{n\times n}$.

(3) 若 $m = n$,则

$$A = PW = WQ,$$

其中 $W \in \mathbb{C}^{n\times n}$ 是酉矩阵,$P, Q \in \mathbb{C}^{n\times n}$ 是半正定的,

$$P^2 = AA^*,\ Q^2 = A^*A.$$

在所有情形中,半正定因子 P 和 Q 都是由 A 唯一确定的,它们的特征值与 A 的奇异值相同.当 $A \in \mathbb{C}^{n\times n}$ 为非奇异时,P 和 Q 是正定的.

证 如果 $n \geq m$,$A = U\Sigma V^*$ 是奇异值分解,其中 $U \in \mathbb{C}^{m\times m}$ 和 $V \in \mathbb{C}^{n\times n}$ 是酉矩阵,而 $\Sigma \in \mathbb{C}^{m\times n}$ 可以写成

$$\Sigma = \begin{bmatrix} S & 0 \end{bmatrix},\ S \equiv \mathrm{diag}(\sigma_1, \cdots, \sigma_m) \in \mathbb{C}^{m\times m}.$$

相应地记

$$V = \begin{bmatrix} V_1 & V_2 \end{bmatrix},\quad V_1 \in \mathbb{C}^{n\times m}, V_2 \in \mathbb{C}^{n\times(n-m)},$$

那么

$$A = U\begin{bmatrix} S & 0 \end{bmatrix}\begin{bmatrix} V_1 & V_2 \end{bmatrix}^* = USV_1^* = (USU^*)(UV_1^*).$$

注意到 $P \equiv USU^*$ 是半正定的,$Y \equiv UV_1^*$ 满足

$$YY^* = UV_1^*V_1U^* = UIU^* = I,$$

(1)得证.

将(1)应用于 A^* 即得(2).

对于(3),注意到

$$A = U\Sigma V^* = (U\Sigma U^*)(UV^*) = (UV^*)(V\Sigma V^*),$$

所以可取 $P = U\Sigma U^*, Q = V\Sigma V^*, W = UV^*$. □

任一复数 z 不仅有加性分解:

$$z = a + \mathrm{i}b,\ a, b \in \mathbb{R},$$

而且有乘性分解:

$$z = \mathrm{e}^{\mathrm{i}\theta}r,\ \left|\mathrm{e}^{\mathrm{i}\theta}\right| = 1,\ r > 0.$$

以上定理中的矩阵分解类似于复数的乘性分解,故称之为**极分解**

(polar decomposition)或**极形式**(polar form).

5.4.12 定义 设 $S=[s_{ij}]\in\mathbb{R}^{n\times n}, S\geq 0$,且满足

$$\sum_{j=1}^{n} s_{ij}\leq 1,\ i=1,\cdots,n;\quad \sum_{i=1}^{n} s_{ij}\leq 1,\ j=1,\cdots,n, \tag{4.27}$$

则称 S 是**次双随机矩阵**(doubly substochastic matrix).

显然,双随机矩阵必是次双随机矩阵.次双随机矩阵有类似于双随机矩阵的特征.

5.4.13 定理 设 $A=[a_{ij}]\in\mathbb{C}^{m\times n}$ 的奇异值分解是 $A=U\Sigma V^*$,其中 $U=[u_{ij}]\in\mathbb{C}^{m\times m}$ 和 $V=[v_{ij}]\in\mathbb{C}^{n\times n}$ 是酉矩阵,$\Sigma=[\sigma_{ij}]\in\mathbb{C}^{m\times n}$ 具有

$$\sigma_{ii}=\sigma_i,\quad i=1,\cdots,q\equiv\min\{m,n\},$$

其余 $\sigma_{ij}=0$,则 A 的主对角元素构成的向量

$$a=\left(a_{11},\cdots,a_{qq}\right)^{\mathrm{T}}\in\mathbb{C}^{q}$$

和 A 的奇异值构成的向量

$$\sigma=\left(\sigma_1,\cdots,\sigma_q\right)\in\mathbb{C}^{q}$$

存在如下线性变换关系:

$$a=S\sigma, \tag{4.28}$$

其中 S 是矩阵 U 和 $\overline{V}$ 的左上 $q\times q$ 主子矩阵的 Hadamard 积(见 **7.2.1**):

$$S=\left[u_{ij}\overline{v_{ij}}\right]\in\mathbb{C}^{q\times q}, \tag{4.29}$$

而且 $|S|$ 是次双随机矩阵.

证 从 $A=U\Sigma V^*$ 直接计算,得

$$a_{ii}=u_{i1}\sigma_1\overline{v_{i1}}+\cdots+u_{iq}\sigma_q\overline{v_{iq}},\ i=1,\cdots,q,$$

此即(4.28)的分量形式.另外,有

$$\left(\sum_{j=1}^{q}\left|u_{ij}\overline{v_{ij}}\right|\right)^2 \le \sum_{j=1}^{q}\left|u_{ij}\right|^2\sum_{j=1}^{q}\left|v_{ij}\right|^2 \le \sum_{j=1}^{m}\left|u_{ij}\right|^2\sum_{j=1}^{n}\left|v_{ij}\right|^2 = 1,\quad i=1,\cdots,q$$

和

$$\left(\sum_{i=1}^{q}\left|u_{ij}\overline{v_{ij}}\right|\right)^2 \le \sum_{i=1}^{q}\left|u_{ij}\right|^2\sum_{i=1}^{q}\left|v_{ij}\right|^2 \le \sum_{i=1}^{m}\left|u_{ij}\right|^2\sum_{i=1}^{n}\left|v_{ij}\right|^2 = 1,\quad j=1,\cdots,q,$$

因此,$|S|$是次双随机矩阵. □

这一定理平行于 **4.5.6**.

5.4.14 定义 $P\in\mathbb{R}^{n\times n}$ 称为**偏置换矩阵**或**部分置换矩阵**(partial permutation matrix),如果 P 的每行和每列全多有一个非零元素,而且这些非零元素全为1.

显然,每个偏置换矩阵是次双随机矩阵,并可通过在一个置换矩阵中用0替换某些1元素而得.而且,置换矩阵的每个方子矩阵是次双随机矩阵.

5.4.15 定理 设 $S\in\mathbb{R}^{n\times n}$,$S\ge 0$,则下列各条件等价:

(1) S 是次双随机矩阵.

(2) S 有双随机扩张,也就是说,S 是一双随机矩阵的左上主子矩阵.

(3) S 是偏置换矩阵的有限凸组合.

(4) 存在双随机矩阵 $\hat{S}\in\mathbb{R}^{n\times n}$,使得 $0\le S\le\hat{S}$.

证 若 $S=[s_{ij}]\in\mathbb{R}^{n\times n}$ 是次双随机矩阵,构造

$$\begin{bmatrix} S & I-D_r \\ I-D_c & S^{\mathrm{T}} \end{bmatrix}\in\mathbb{R}^{2n\times 2n}, \tag{4.30}$$

其中 D_r 与 D_c 分别是 S 的行和的对角矩阵

$$D_r \equiv \mathrm{diag}(s_{11}+\cdots+s_{1n},\cdots,s_{n1}+\cdots+s_{nn})$$

与列和的对角矩阵

$$D_c \equiv \mathrm{diag}\left(s_{11}+\cdots+s_{n1},\cdots,s_{1n}+\cdots+s_{nn}\right).$$

容易验证(4.30)是双随机矩阵,它是S的双随机扩张.

依 **4.5.4**,双随机矩阵(4.30)是$2n\times 2n$置换矩阵的凸组合.显然,这些置换矩阵左上$n\times n$主子矩阵的同一凸组合就是S;而且每个如此主子矩阵是偏置换矩阵.因此,每个次双随机矩阵是偏置换矩阵的有限凸组合.反之,偏置换矩阵的有限凸组合形如

$$P\equiv[p_{ij}]=\sum_{k=1}^{m}\alpha_k P_k;\ \sum_{k=1}^{m}\alpha_k=1;\alpha_k\ge 0,k=1,\cdots,m,$$

其中$P_k=\left[p_{ij}^{(k)}\right]\in\mathbb{R}^{n\times n}$是偏置换矩阵,$k=1,\cdots,m$;由于有

$$\sum_{j=1}^{n}p_{ij}=\sum_{j=1}^{n}\sum_{k=1}^{m}\alpha_k p_{ij}^{(k)}=\sum_{k=1}^{m}\alpha_k\sum_{j=1}^{n}p_{ij}^{(k)}\le\sum_{k=1}^{m}\alpha_k\le 1,$$
$$i=1,\cdots,n$$

和

$$\sum_{i=1}^{n}p_{ij}=\sum_{i=1}^{n}\sum_{k=1}^{m}\alpha_k p_{ij}^{(k)}=\sum_{k=1}^{m}\alpha_k\sum_{i=1}^{n}p_{ij}^{(k)}\le\sum_{k=1}^{m}\alpha_k\le 1,$$
$$j=1,\cdots,n,$$

从而P是次双随机矩阵.

最后,如果将给定的次双随机矩阵S表示为偏置换矩阵的凸组合;在其每个偏置换矩阵中,用1替换某些零元素,使之成为置换矩阵.那么得到一个置换矩阵的凸组合$\hat{S}$,它是双随机矩阵,而且$\hat{S}\ge S$.反之,若有双随机矩阵$\hat{S}\in\mathbb{R}^{n\times n}$,使得$0\le S\le\hat{S}$,则$S$显然是次双随机矩阵. □

5.4.16 定理 设

$$x=\left(x_1,\cdots,x_n\right)^{\mathrm{T}},y=\left(y_1,\cdots,y_n\right)^{\mathrm{T}}\in\mathbb{R}^n,x\ge 0,y\ge 0,$$

则 y 弱优化于 x 的充分必要条件是存在次双随机矩阵 $S \in \mathbb{R}^{n\times n}$,使得 $x = Sy$.

证 设 $S = [s_{ij}] \in \mathbb{R}^{n\times n}$ 是次双随机矩阵,使得 $x = Sy$.

依 **5.4.15**,存在双随机矩阵 $\hat{S} \in \mathbb{R}^{n\times n}$,使得

$$0 \le S \le \hat{S}.$$

采用 **4.5.7** 中的记号表示实向量按代数递减顺序重新编号,并设

$$(Sy)_{i_j} = (Sy)_{[j]}, \quad j = 1, \cdots, n,$$

则

$$\sum_{i=1}^{k} x_{[i]} = \sum_{i=1}^{k} (Sy)_{[i]} = \sum_{j=1}^{k} (Sy)_{i_j}$$

$$\le \sum_{j=1}^{k} (\hat{S}y)_{i_j} \le \sum_{i=1}^{k} (\hat{S}y)_{[i]} \le \sum_{i=1}^{k} y_{[i]}, k = 1, \cdots, n, \quad (4.31)$$

其中最后的不等号依据 **4.5.8** 中的推导.因此 y 弱优化于 $Sy = x$.

反之,假设 y 弱优化于 x,分 $x = 0$ 和 $x \ne 0$ 两种情况:

若 $x = 0$,直接取 $S = 0$.

若 $x \ne 0$,用 ε 表示 x 的最小正元素,令

$$\delta \equiv (y_1 - x_1) + \cdots + (y_n - x_n),$$

显然,$\delta \ge 0$.取 m 是使得 $\varepsilon \ge \delta / m$ 的任一正整数.设

$$\xi \equiv (x_1, \cdots, x_n, \delta / m, \cdots, \delta / m)^{\mathrm{T}} \in \mathbb{R}^{n+m}$$

和

$$\eta \equiv (y_1, \cdots, y_n, 0, \cdots, 0) \in \mathbb{R}^{n+m},$$

于是有

$$\sum_{i=1}^{k} \xi_{[i]} \le \sum_{i=1}^{k} \eta_{[i]}, \quad k = 1, \cdots, n+m,$$

而且当 $k = n + m$ 时成立等号.因此,η 强优化于 ξ.依 **4.5.8**,存在双随机矩阵

$$\hat{S} \in \mathbb{R}^{(n+m)\times(n+m)}$$

使得 $\xi = \hat{S}\eta$.现在用 S 表示 $\hat{S}$ 的左上 $n \times n$ 主子矩阵,显然 S 是次双随机矩阵而且 $x = Sy$. □

此定理平行于 **4.5.8.**

5.4.17 推论 设

$$x = (x_1, \cdots, x_n)^{\mathrm{T}}, y = (y_1, \cdots, y_n)^{\mathrm{T}} \in \mathbb{R}^n,$$

则 y 弱优化于 x 的充分必要条件是存在双随机矩阵 $S \in \mathbb{R}^{n\times n}$,使得 $x \le Sy$.

证 若存在双随机矩阵 $S \in \mathbb{R}^{n\times n}$,使得 $x \le Sy$,则借鉴(4.31),有

$$\sum_{i=1}^{k} x_{[i]} \le \sum_{i=1}^{k} (Sy)_{[i]} \le \sum_{i=1}^{k} y_{[i]}, \quad k = 1, \cdots, n,$$

因此 y 弱优化于 x .

反之,假设 y 弱优化于 x .令 $e = (1, \cdots, 1)^{\mathrm{T}} \in \mathbb{R}^n$,取 $\kappa > 0$ 足够大,使得

$$x + \kappa e > 0, \quad y + \kappa e > 0.$$

现在,因为 $y + \kappa e$ 弱优化于 $x + \kappa e$,依 **5.4.16,**存在次双随机矩阵 $\hat{S}$,使得

$$x + \kappa e = \hat{S}(y + \kappa e).$$

再依 **5.4.15,**存在双随机矩阵 $S \ge \hat{S}$,因此

$$x + \kappa e = \hat{S}(y + \kappa e) \le S(y + \kappa e) = Sy + \kappa e,$$

从而 $x \le Sy$. □

5.4.18 定义 将置换矩阵中各个1元素替换为绝对值为1的实或复元素而得到的矩阵,称为**广义置换矩阵**(generalized permutation matrix).

5.4.19 引理 每个偏置换矩阵是两个实广义置换矩阵的凸组合.

证 设 $P=[p_{ij}]\in\mathbb{R}^{n\times n}$ 是偏置换矩阵,依 **5.4.14**, P 的每行和每列至多有一个等于1的非零元素.因此,如果 P 的第 $i_1,\cdots,i_m$ 行的元素全为零,则必存在第 $j_1,\cdots,j_m$ 列,它们的元素也全为零.定义矩阵 $G_1=\left[g_{ij}^{(1)}\right],G_2=\left[g_{ij}^{(2)}\right]\in\mathbb{R}^{n\times n}$:

$$\begin{cases} g_{i_kj_k}^{(1)}=-g_{i_kj_k}^{(2)}=1, & k=1,\cdots,m, \\ g_{ij}^{(1)}=g_{ij}^{(2)}=p_{ij}, & \text{其余}. \end{cases}$$

显然, G_1 是置换矩阵, G_2 是广义置换矩阵,而且

$$P=\frac{1}{2}G_1+\frac{1}{2}G_2,$$

这表明 P 是 G_1 和 G_2 的凸组合. □

5.4.20 引理 设 $G\in\mathbb{C}^{n\times n}$ 是广义置换矩阵,并设

$$\sigma=(\sigma_1,\cdots,\sigma_n)^{\mathrm{T}}\in\mathbb{R}^n,\sigma\geq 0,$$

记

$$G\sigma=(\hat{\sigma}_1,\ \cdots,\ \hat{\sigma}_n)^{\mathrm{T}},$$

则 $\operatorname{diag}(\hat{\sigma}_1,\cdots,\hat{\sigma}_n)$ 的奇异值是 $\sigma_1,\cdots,\sigma_n$.

证明留作练习.

5.5 Householder 变换

5.5.1 定义 设 $w=(w_1,\cdots,w_n)^{\mathrm{T}}\in\mathbb{R}^n,\|w\|_2=1$,矩阵

$$H = I - 2ww^{\mathrm{T}}$$

$$= \begin{bmatrix} 1-2w_1w_1 & -2w_1w_2 & \cdots & -2w_1w_n \\ -2w_2w_1 & 1-2w_2w_2 & \cdots & -2w_2w_n \\ \cdots & \cdots & \cdots & \cdots \\ -2w_nw_1 & -2w_nw_2 & \cdots & 1-2w_nw_n \end{bmatrix} \tag{5.1}$$

称为相伴 w 的 **Householder 变换**或 **Householder 矩阵**或**镜像变换**.

显然,ww^{T} 是一个秩 1 矩阵(即秩为 1 的矩阵).因此,可以说,Householder 变换 H 是对单位矩阵作秩 1 校正的矩阵.

5.5.2 定理 设 H 是 Householder 变换,形如(5.1),则

(1) H 是实对称的正交矩阵.

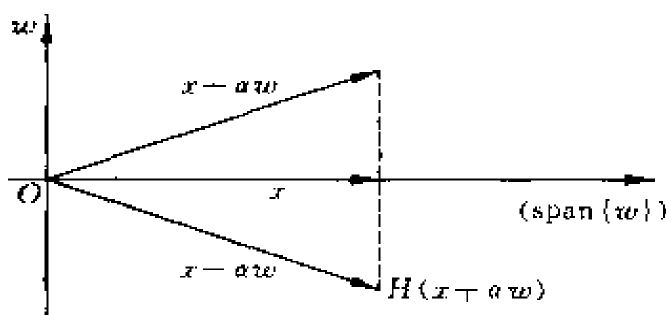
(2) 1 是 H 的 $n-1$ 重特征值;-1 是 H 的单特征值,且 w 是其相应的单位特征向量.

(3) $\det H = -1$.

(4) 成立

$$H(x+\alpha w) = x - \alpha w,\ \ \forall x \in (\mathrm{span}\{w\})^{\perp}, \alpha \in \mathbb{R}. \tag{5.2}$$

几何上看,H 是关于与 w 正交的过原点超平面的**反射变换**(见图):



证 (1) H 显然是实对称的,而且容易验证 $H^{\mathrm{T}}H = I$.

(2) 因 $\mathrm{rank}(ww^{\mathrm{T}}) = 1$,故

$$\det(I - H) = 2\det(ww^{\mathrm{T}}) = 0.$$

而且线性方程组

$$(I-H)x=0$$

有 $n-1$ 个线性无关的解,这表明1是 H 的代数重数和几何重数同为 $n-1$ 的特征值.另外,

$$Hw=w-2ww^{\mathrm{T}}w=w-2w=-w$$

表明 -1 是 H 的1重特征值, w 是相应的特征向量.

(3) 是(2)的推论: $\det H=1^{n-1}\cdot(-1)=-1$.

(4) $\forall x\in(\mathrm{span}\{w\})^{\perp}$,有 $w^{\mathrm{T}}x=0$,因此

$$\begin{aligned}H(x+\alpha w)&=x+\alpha w-2ww^{\mathrm{T}}(x+\alpha w)\\&=x+\alpha w-2ww^{\mathrm{T}}x-2\alpha w\\&=x-\alpha w,\quad \alpha\in\mathbb{R}.\end{aligned}$$

□

5.5.3 定理 对于任一 $x\in\mathbb{R}^n$, $x\neq 0$,可构造单位向量 $w\in\mathbb{R}^n$,使得相伴 w 的 Householder 变换 H 满足

$$Hx=\pm\|x\|_2 e_1, \tag{5.3}$$

这里 $e_1=(1,0,\cdots,0)^{\mathrm{T}}\in\mathbb{R}^n$.

(5.3)是说, H 将 x 的后 $n-1$ 个分量约化为零.

证 注意到 $w^{\mathrm{T}}x$ 是实数,有

$$Hx=(I-2ww^{\mathrm{T}})x=x-2(w^{\mathrm{T}}x)w$$

于是,欲使(5.3)成立,必须

$$2(w^{\mathrm{T}}x)w=x\pm\|x\|_2 e_1. \tag{5.4}$$

假定 $x\pm\|x\|_2 e_1\neq 0$(否则 x 的后 $n-1$ 个分量已经为零),且选取单位向量 w 使 $w^{\mathrm{T}}x>0$,则由(5.4)有

$$2(w^{\mathrm{T}}x)=\left\|2(w^{\mathrm{T}}x)w\right\|_2=\left\|x\pm(\|x\|_2)e_1\right\|_2.$$

将此式代入(5.4),取

$$w=\frac{x\pm\|x\|_2 e_1}{\left\|x\pm(\|x\|_2)e_1\right\|_2}, \tag{5.5}$$

即能使(5.3)成立. □

这一定理是Householder变换最具特征的性质,而且其证明提供了构造 w 的方案.

注意,实际应用时公式(5.5)中的正负号选取应与所要约化的 $x=\left(x_1,\cdots,x_n\right)^{\mathrm{T}}$ 的第1个分量 x_1 的正负号保持一致,即

$$w=\frac{x+\left(\operatorname{sgn} x_1\right)\|x\|_2 e_1}{\left\|x+\left(\left(\operatorname{sgn} x_1\right)\|x\|_2\right) e_1\right\|_2}, \tag{5.6}$$

这样,可以避免出现大的计算误差.

5.6 Hessenberg 矩阵

5.6.1 定义 $H=[h_{ij}]\in\mathbb{C}^{n\times n}$ 称为上 **Hessenberg 矩阵**,如果

$$h_{ij}=0,\quad \forall i\geq j+2. \tag{6.1}$$

也就是说, H 形如

$$H=\begin{bmatrix} h_{11} & h_{12} & \cdots & \cdots & h_{1n} \\ h_{21} & h_{22} & \ddots & & \vdots \\ 0 & \ddots & \ddots & \ddots & \vdots \\ \vdots & \ddots & \ddots & \ddots & h_{n-1,n} \\ 0 & \cdots & 0 & h_{n,n-1} & h_{nn} \end{bmatrix}. \tag{6.2}$$

当 H 的次对角元素

$$h_{i+1,i}\neq 0,\ i=1,\cdots,n-1 \tag{6.3}$$

时称其为**不可约**的.

类似地，H 称为**下 Hessenberg 矩阵**，如果

$$h_{ij}=0, \quad \forall j \geq i+2.$$

Hessenberg 矩阵在零元素所占比例及分布上接近三角矩阵.虽然在特征值等性质方面,Hessenberg 矩阵远不如三角矩阵那样简单;但是从应用相似变换看,一般,将一个矩阵约化成 Hessenberg 矩阵是可行的,约化成三角矩阵则不易实现;而且通过约化成 Hessenberg 形来处理矩阵计算(主要是迭代计算)常能大大节省运算量.

5.6.2 定理 设 $A=[a_{ij}]\in\mathbb{R}^{n\times n}$,则存在 $n\times n$ 正交矩阵 P,使得

$$P^{-1}AP=H, \tag{6.4}$$

其中 H 是形如(6.2)的上 Hessenberg 矩阵.完成如此约化的运算量为 $O(n^3)$.

证 利用数学归纳法. $n\leq 2$ 时结论显然成立.

现在考虑 $n\geq 3$. 假设结论对所有 $n-1$ 阶矩阵成立.应用 Householder 变换 H_1 约化 A 的第1列,由于进行相似变换 H_1AH_1,为了保证 A 左乘 H_1 后的第1列已专门约化为零的元素不会在右乘 H_1 后重新成为非零,应取 H_1 形如

$$H_1=\begin{bmatrix}1 & 0\\ 0 & \tilde{H}_1\end{bmatrix}, \tag{6.5}$$

其中 $\tilde{H}_1$ 是 $n-1$ 阶 Householder 矩阵. H_1 相当于(5.1)中取 w 的第1个分量 $w_1=0$ 的情形.将 A 写成相应的分块形式

$$A=\begin{bmatrix}h_{11} & \left(b^{(1)}\right)^{\mathrm{T}}\\ a^{(1)} & \tilde{A}\end{bmatrix}, \tag{6.6}$$

其中

$$h_{11}=a_{11}, a^{(1)}=\left(a_{21},\cdots,a_{n1}\right)^{\mathrm{T}}, b^{(1)}=\left(a_{12},\cdots,a_{1n}\right)^{\mathrm{T}},$$

$\tilde{A}$ 是 $n-1$ 阶子矩阵.于是

$$H_1AH_1=\begin{bmatrix} h_{11} & \left(b^{(1)}\right)^{\mathrm{T}}\tilde{H}_1 \\ \tilde{H}_1a^{(1)} & \tilde{H}_1\tilde{A}\tilde{H}_1 \end{bmatrix}. \tag{6.7}$$

依 5.5.3,可构造 $\tilde{H}_1$ 使得

$$\tilde{H}_1a^{(1)}=(\operatorname{sgn}a_{21})\left\|a^{(1)}\right\|_2e_1, \tag{6.8}$$

这里 $e_1=(1,0,\cdots,0)^{\mathrm{T}}$ 是 $n-1$ 维单位向量.因此

$$H_1AH_1=\left[\begin{array}{c|c} h_{11} & \left(b^{(1)}\right)^{\mathrm{T}}\tilde{H}_1 \\ \hline h_{21} & \\ 0 & \\ \vdots & \tilde{H}_1\tilde{A}\tilde{H}_1 \\ 0 & \end{array}\right], \tag{6.9}$$

其中

$$h_{21}\equiv(\operatorname{sgn}a_{21})\left\|a^{(1)}\right\|_2. \tag{6.10}$$

根据归纳法假设,存在 $n-1$ 阶正交矩阵 $\tilde{P}$,使得

$$\tilde{P}^{-1}\left(\tilde{H}_1\tilde{A}\tilde{H}_1\right)\tilde{P}=\tilde{H},$$

$\tilde{H}$ 是 $n-1$ 阶上 Hessenberg 矩阵.

基于(6.5)取法的同样原因,即知 $\tilde{P}$ 形如

$$\tilde{P}=\begin{bmatrix} 1 & 0 \\ 0 & \hat{P} \end{bmatrix},$$

其中 $\hat{P}$ 是 $n-2$ 阶正交矩阵.令

$$P=H_1\begin{bmatrix} 1 & 0 \\ 0 & \tilde{P} \end{bmatrix},$$

则

$$P^{-1}AP=\begin{bmatrix} 1 & 0 \\ 0 & \tilde{P}^{-1} \end{bmatrix}H_1AH_1\begin{bmatrix} 1 & 0 \\ 0 & \tilde{P} \end{bmatrix},$$

$$
=\left[\begin{array}{c|c}
h_{11} & \left(b^{(1)}\right)^{\mathrm{T}} \tilde{H}_1 \tilde{P} \\
\hline
h_{21} & \\
0 & \\
\vdots & \tilde{H} \\
0 &
\end{array}\right] \equiv H,
$$

H 是上 Hessenberg 矩阵.约化 A 为 Hessenberg 形的实用算法及其运算量可以在许多数值分析的著作中找到. □

以上定理的证明也提供了将 A 约化成 Hessenberg 矩阵的方案.事实上,在得到(6.9)后,若仿照(6.6),令

$$
A_1 = \tilde{H}_1 \tilde{A} \tilde{H}_1 = \begin{bmatrix} h_{22} & \left(b^{(2)}\right)^{\mathrm{T}} \\ a^{(2)} & \tilde{A}_1 \end{bmatrix},
$$

这是 $n-1$ 阶子矩阵,则存在 $n-1$ 阶矩阵

$$
\hat{H}_2 = \begin{bmatrix} 1 & 0 \\ 0 & \tilde{H}_2 \end{bmatrix},
$$

这里 $\tilde{H}_2$ 是 $n-2$ 阶 Householder 矩阵,使得

$$
\hat{H}_2 A_1 \hat{H}_2 = \left[\begin{array}{c|c}
h_{22} & \left(b^{(2)}\right)^{T} \tilde{H}_2 \\
\hline
h_{32} & \\
0 & \\
\vdots & \tilde{H}_2 \tilde{A}_1 \tilde{H}_2 \\
0 &
\end{array}\right],
$$

其中

$$
h_{32} \equiv \left(\operatorname{sgn} a_{32}^{(2)}\right) \left\| a^{(2)} \right\|_2,
$$

$a_{32}^{(2)}$ 是 $n-2$ 维向量 $a^{(2)}$ 的第1个分量.于是,令

$$H_2=\begin{bmatrix}1 & 0\\ 0 & \hat{H}_2\end{bmatrix}$$

便有

$$H_2H_1AH_1H_2=\begin{bmatrix}h_{11} & * & * & \cdots & *\\ h_{21} & h_{22} & * & \cdots & *\\ 0 & h_{32} & h_{33} & \cdots & *\\ 0 & 0 & * & \cdots & *\\ \vdots & \vdots & \vdots & \vdots & \vdots\\ 0 & 0 & * & \cdots & *\end{bmatrix}.$$

如此继续，一共 $n-2$ 次,可约化成上 Hessenberg 矩阵 H：

$$H_{n-2}\cdots H_2H_1AH_1H_2\cdots H_{n-2}=H. \tag{6.11}$$

最后,令

$$P=H_1H_2\cdots H_{n-2},$$

P 是正交矩阵,即有 $P^{-1}AP=H$.

一般, Hessenberg 分解是不唯一的,但是下一定理表明此类分解仍具有很强的确定性.

5.6.3 定理 设 $A\in\mathbb{R}^{n\times n}$ 有两个上 Hessenberg 分解:

$$P^{\mathrm{T}}AP=H,\quad Q^{\mathrm{T}}AQ=G, \tag{6.12}$$

其中

$$P=\begin{bmatrix}p^{(1)} & \cdots & p^{(n)}\end{bmatrix} \text{ 和 } Q=\begin{bmatrix}q^{(1)} & \cdots & q^{(n)}\end{bmatrix}$$

是 n 阶正交矩阵;而

$$H\equiv[h_{ij}]=\begin{bmatrix}h^{(1)} & \cdots & h^{(n)}\end{bmatrix}$$

和

$$G\equiv[g_{ij}]=\begin{bmatrix}g^{(1)} & \cdots & g^{(n)}\end{bmatrix}$$

是上 Hessenberg 矩阵.

进而,设 H 是不可约的,而且 $p^{(1)}=q^{(1)}$,则存在对角元素或1或 -1 的对角矩阵 D,使得

$$P = DQ,\ H = DGD. \tag{6.13}$$

这蕴涵:

$$p^{(i)} = \pm q^{(i)}, i = 1,\cdots,n,$$

而且

$$\left|h_{ij}\right| = \left|g_{ij}\right|,\ \ i,j = 1,\cdots,n.$$

证 利用数学归纳法.

已设 $p^{(1)} = q^{(1)}$.

现假定对于某个 $m\ (1 \le m < n)$成立

$$p^{(i)} = \varepsilon_i q^{(i)},\ \ i = 1,\cdots,m, \tag{6.14}$$

其中 $\varepsilon_i = 1$ 或 -1.要证明存在 $\varepsilon_{m+1} = 1$ 或 -1,使得

$$p^{(m+1)} = \varepsilon_{m+1} q^{(m+1)}. \tag{6.15}$$

从(6.12),有

$$\begin{aligned} Ap^{(m)} &= Ph^{(m)} \\ &= h_{1m}p^{(1)} + \cdots + h_{mm}p^{(m)} + h_{m+1,m}p^{(m+1)} \end{aligned} \tag{6.16}$$

和

$$\begin{aligned} Aq^{(m)} &= Qg^{(m)} \\ &= g_{1m}q^{(1)} + \cdots + g_{mm}q^{(m)} + g_{m+1,m}q^{(m+1)}, \end{aligned} \tag{6.17}$$

于是

$$h_{im} = \left(p^{(i)}\right)^T Ap^{(m)},\quad i = 1,\cdots,m$$

和

$$g_{im} = \left(q^{(i)}\right)^T Aq^{(m)},\quad i = 1,\cdots,m,$$

由此及(6.14),

$$h_{im} = \left(\varepsilon_i q^{(i)}\right)^T A\left(\varepsilon_m q^{(m)}\right) = \varepsilon_i \varepsilon_m g_{im},\quad i = 1,\cdots,m,$$

将这些关系式代入(6.16),并利用(6.14)和(6.17),得

$$\begin{aligned} h_{m+1,m}p^{(m+1)} &= \varepsilon_m\left(Aq^{(m)} - \varepsilon_1^2 g_{1m}q^{(1)} - \cdots - \varepsilon_m^2 g_{mm}q^{(m)}\right) \\ &= \varepsilon_m\left(Aq^{(m)} - g_{1m}q^{(1)} - \cdots - g_{mm}q^{(m)}\right) \end{aligned}$$

$$= \varepsilon_m g_{m+1,m} q^{(m+1)}. \tag{6.18}$$

根据此式,

$$|h_{m+1,m}| = \|h_{m+1,m} p^{(m+1)}\|_2 = \|\varepsilon_m g_{m+1,m} q^{(m+1)}\|_2 = |g_{m+1,m}|.$$

又因 H 不可约, $h_{m+1,m} \neq 0$,故

$$g_{m+1,m}/h_{m+1,m} = 1 \text{ 或 } -1,$$

因此(6.18)可写成(6.15)的形式. □

这一定理表明,对于 $P^{T}AP = H$ 是不可约 Hessenberg 矩阵的情况而言,在可以相差正负号的意义下,P 和 H 完全由 P 的第1列确定.

5.6.4 定理 设 $H = [h_{ij}] \in \mathbb{C}^{n\times n}$ 是不可约上 Hessenberg 矩阵, $x_1, \cdots, x_{n-1}$ 满足

$$\begin{bmatrix} h_{21} & h_{22}-\lambda & h_{23} & \cdots & h_{2,n-1} \\ 0 & h_{32} & \ddots & \ddots & \vdots \\ \vdots & \ddots & \ddots & \ddots & h_{n-2,n-1} \\ \vdots & & \ddots & \ddots & h_{n-1,n-1}-\lambda \\ 0 & \cdots & \cdots & 0 & h_{n,n-1} \end{bmatrix} \begin{bmatrix} x_1 \\ \vdots \\ \vdots \\ x_{n-2} \\ x_{n-1} \end{bmatrix} + \begin{bmatrix} h_{2n} \\ \vdots \\ \vdots \\ h_{n-1,n} \\ h_{nn}-\lambda \end{bmatrix} = 0. \tag{6.19}$$

并记

$$k(\lambda) = (h_{11}-\lambda)x_1 + h_{12}x_2 + \cdots + h_{1,n-1}x_{n-1} + h_{1n}, \tag{6.20}$$

则 H 的特征多项式

$$\det(H-\lambda I) = (-1)^{n+1} k(\lambda) h_{21} h_{32} \cdots h_{n,n-1}. \tag{6.21}$$

证 令

$$x = (x_1, \cdots, x_{n-1}, 1),$$

联立(6.19)和(6.20),可得

$$(H-\lambda I)x = k(\lambda)e_1, \tag{6.22}$$

其中 $e_1 = (1,0,\cdots,0)^{T}$ 是 n 维单位向量.

等式(6.22)的意思是：矩阵 $H-\lambda I$ 的前 $n-1$ 列分别以 $x_1,\cdots,x_{n-1}$ 为系数的线性组合加到最后第 n 列等于向量 $k(\lambda)e_1$.

因此,若将 $H-\lambda I$ 的第 n 列替换为 $k(\lambda)e_1$ 并记之为 G，则根据行列式的基本性质,

$$\det(H-\lambda I)=\det G$$

$$=\begin{bmatrix} h_{11}-\lambda & h_{12} & \cdots & h_{1,n-1} & k(\lambda) \\ h_{21} & h_{22}-\lambda & \cdots & h_{2,n-1} & 0 \\ 0 & h_{32} & \ddots & \vdots & \vdots \\ \vdots & \ddots & \ddots & h_{n-1,n-1}-\lambda & 0 \\ 0 & \cdots & 0 & h_{n,n-1} & 0 \end{bmatrix}$$

$$=(-1)^{n+1}k(\lambda)h_{21}h_{32}\cdots h_{n,n-1}.$$ □

这一定理提供了求不可约上 Hessenberg 矩阵 H 的特征多项式 $\det(H-\lambda I)$ 之值的一个方案:

(1) 解一个三角形方程组(6.19),递推地产生 $x_{n-1},\cdots,x_1$;

(2) 从(6.20)算出 $k(\lambda)$;

(3) 由(6.21)得出 $\det(H-\lambda I)$.

次对角元素 $h_{21},h_{32},\cdots,h_{n,n-1}$ 起着特殊作用.

由于 $\det(H-\lambda I)$ 和 $k(\lambda)$ 仅差常数因子,因此计算 H 的特征值也可转而求 $k(\lambda)$ 的零点.

5.7 Givens 变换和 QR 分解

5.7.1 定义　Givens 变换是指形如

$$G(i,k,\theta)=\begin{bmatrix}1 & 0 & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & 0\\ 0 & \ddots & \ddots & & & & & & & & \vdots\\ \vdots & \ddots & 1 & 0 & \cdots & \cdots & \cdots & \cdots & 0 & & \vdots\\ \vdots & & 0 & c & 0 & \cdots & 0 & s & \vdots & & \vdots\\ \vdots & & \vdots & 0 & 1 & \ddots & & 0 & \vdots & & \vdots\\ \vdots & & \vdots & \vdots & \ddots & \ddots & \ddots & \vdots & \vdots & & \vdots\\ \vdots & & \vdots & 0 & & \ddots & 1 & 0 & \vdots & & \vdots\\ \vdots & & \vdots & -s & 0 & \cdots & 0 & c & 0 & & \vdots\\ \vdots & & 0 & \cdots & \cdots & \cdots & \cdots & 0 & 1 & \ddots & \vdots\\ \vdots & & & & & & & & \ddots & \ddots & 0\\ 0 & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & 0 & 1\end{bmatrix}\begin{matrix}\\ \\ \\ i\\ \\ \\ \\ k\\ \\ \\ \\ \end{matrix}$$

$$\qquad\qquad\qquad i \qquad\qquad\qquad k$$

$$(7.1)$$

的 $n\times n$ 矩阵,其中 $c=\cos\theta$, $s=\sin\theta$, $\theta\in\mathbb{R}, i\neq k$.

几何上,对于任何向量 $x\in\mathbb{R}^n$, $G(i,k,\theta)x$ 等于 x 在 (i,k) 坐标平面内旋转 θ 度后所得的向量.因此,Givens 变换也称为**平面旋转变换**.

显然, $G(i,k,\theta)$ 是一个正交矩阵,因为

$$\begin{vmatrix}c & s\\ -s & c\end{vmatrix}=c^2+s^2=1,$$

所以,可以说, $G(i,k,\theta)$ 是对单位矩阵作秩 2 校正的矩阵.

Householder 变换对于大比例的约化元素为零是非常有效的,例如将一个向量除其第 1 个分量之外全部零化.然而,在许多计算中必须对要零化元素加以更细的选择.Givens 变换正是这样一种工具.

5.7.2 定理 设$x=\left(x_1,\cdots,x_n\right)^{\mathrm{T}}\in\mathbb{R}^n$,而且

$$y=\left(y_1,\cdots,y_n\right)^{\mathrm{T}}=G\left(i,k,\theta\right)x\,, \tag{7.2}$$

其中$G\left(i,k,\theta\right)$形如(6.1), 则取

$$c=\frac{x_i}{\sqrt{x_i^2+x_k^2}},\quad s=\frac{x_k}{\sqrt{x_i^2+x_k^2}} \tag{7.3}$$

时,$y_k=0$.

证 由(7.2),有

$$y_i=cx_i+sx_k\,,$$

$$y_k=-sx_i+cx_k\,,$$

$$y_j=x_j\,,\quad j\neq i\text{或}k\,.$$

利用(7.3)即得$y_k=0$. □

由于选择x_i有任意性,因此将x的第k个分量约化为零的Givens变换一般是不唯一的.

5.7.3 定义 设$A\in\mathbb{R}^{m\times n}\left(m\geq n\right),Q\in\mathbb{R}^{m\times m}$是正交矩阵,使得

$$A=QR\,, \tag{7.4}$$

其中R是$m\times n$矩阵,形如

$$R=\begin{bmatrix}R_1\\0\end{bmatrix}, \tag{7.5}$$

R_1是$n\times n$上三角矩阵,则称(7.4)是A的**QR分解**.

QR分解在求解最小二乘问题等课题中有重要的应用;尤其在非常突出的计算矩阵特征值问题的QR方法中实现QR分解是其关键的步骤.

5.7.4 定理 设$A\in\mathbb{R}^{m\times n}\left(m\geq n\right),b\in\mathbb{R}^m$,(7.4)是$A$的QR分解,而且

$$Q^{\mathrm{T}}b=\begin{bmatrix}c\\d\end{bmatrix},\tag{7.6}$$

其中$c\in\mathbb{R}^n,d\in\mathbb{R}^{m-n}$,则

$$\min_x\|Ax-b\|_2=\|d\|_2.\tag{7.7}$$

而且此极小值在$x\in\mathbb{R}^n$满足

$$R_1x=c\tag{7.8}$$

时达到.

证 根据假设,

$$\|Ax-b\|_2^2=\left\|Q^{\mathrm{T}}Ax-Q^{\mathrm{T}}b\right\|_2^2=\|R_1x-c\|_2^2+\|d\|_2^2,$$

由此即得定理的结论. □

这一定理表明了怎样利用 QR 分解获得基本的满秩(即$\operatorname{rank}A=n$)最小二乘问题的解:

(1) 实现A的 QR 分解(7.4);

(2) 按(7.6)计算c和d;

(3) 解上三角形方程组(7.8),所得x即最小二乘解;

(4) 计算最小二乘剩余$\|d\|_2$.

实现 QR 分解(7.4)可用 Householder 方法,也可用 Givens 方法,还有改进的 Gram-Schmidt 方法.

5.7.5 定理 设$A\in\mathbb{R}^{m\times n}\ (m\geq n)$,则利用 Householder 变换和 Givens 变换均可产生正交矩阵$Q\in\mathbb{R}^{m\times m}$,使得

$$Q^{\mathrm{T}}A=R\tag{7.9}$$

是形如(7.5)的上三角形.而且

(1) Householder 方法可在

$$n^2(m-n/3)$$

次运算内计算出 $\boldsymbol{R}$.

(2) Givens 方法可在

$$2n^2(m-n/3)$$

次运算内计算出 $\boldsymbol{R}$.

证 依 **5.5.3**,可相继构造 n 个 Householder 矩阵

$$H_1, H_2, \cdots, H_n,$$

使得

H_1A 的第1列的第2个至第 m 个元素为零;

H_2H_1A 继之增添第2列的第3个至第 m 个元素为零;

…

$H_n\cdots H_2H_1A$ 继之增添第 n 列的第 $n+1$ 个至第 m 个元素为零。

从而得到

$$H_n\cdots H_2H_1A \equiv R. \tag{7.10}$$

由此,令

$$Q = H_1H_2\cdots H_n, \tag{7.11}$$

则(7.10)即表现为(7.9).

不难统计,计算 R 可在 $n^2(m-n/3)$ 次运算内完成.

约化的具体过程如下:

$$\xrightarrow{H_1A}\begin{bmatrix} * & * & \cdots & \cdots & * \\ 0 & * & & & * \\ \vdots & \vdots & & & \vdots \\ \vdots & \vdots & & & \vdots \\ \vdots & \vdots & & & \vdots \\ \vdots & \vdots & & & \vdots \\ \vdots & \vdots & & & \vdots \\ 0 & * & \cdots & \cdots & * \end{bmatrix} \xrightarrow{H_2H_1A}\begin{bmatrix} * & * & * & \cdots & * \\ 0 & * & * & & * \\ 0 & 0 & * & & * \\ \vdots & \vdots & \vdots & & \vdots \\ \vdots & \vdots & \vdots & & \vdots \\ \vdots & \vdots & \vdots & & \vdots \\ \vdots & \vdots & \vdots & & \vdots \\ 0 & 0 & * & \cdots & * \end{bmatrix}$$

$$\cdots \xrightarrow{H_n \cdots H_1 A} \begin{bmatrix} * & * & \cdots & * & * \\ 0 & * & \cdots & * & * \\ \vdots & \ddots & \ddots & \vdots & \vdots \\ \vdots & & \ddots & * & * \\ \vdots & & & 0 & * \\ \vdots & & & 0 & 0 \\ \vdots & & & \vdots & \vdots \\ 0 & 0 & \cdots & 0 & 0 \end{bmatrix}.$$

现在考虑利用 Givens 变换.从 A 出发,可相继选取如下 $(m+n-1)n/2$ 个 Givens 变换:

$$G_{ik} \equiv G_{ik}(i,k,\theta_{ik}),\ i<k,\ k=2,\cdots,m,i=1,\cdots,n. \tag{7.12}$$

并令

$$\begin{aligned} Q^{\mathrm{T}} = {} & G_{nm}G_{n-1,m}\cdots G_{1m}\cdots G_{n,n+1}G_{n-1,n+1}\cdots G_{1,n+1} \\ & G_{n-1,n}G_{n-2,n}\cdots G_{1n}\cdots G_{34}G_{24}G_{14}G_{23}G_{13}G_{12} \end{aligned} \tag{7.13}$$

使得成立(7.9).约化的具体过程如下:

$$\xrightarrow{G_{12}A} \begin{bmatrix} * & * & \cdots & \cdots & * \\ 0 & * & & & * \\ * & * & & & * \\ \vdots & \vdots & & & \vdots \\ \vdots & \vdots & & & \vdots \\ \vdots & \vdots & & & \vdots \\ \vdots & \vdots & & & \vdots \\ \vdots & \vdots & & & \vdots \\ * & * & \cdots & \cdots & * \end{bmatrix} \xrightarrow{G_{13}G_{12}A} \begin{bmatrix} * & * & \cdots & \cdots & * \\ 0 & * & & & * \\ 0 & * & & & * \\ * & * & & & * \\ \vdots & \vdots & & & \vdots \\ \vdots & \vdots & & & \vdots \\ \vdots & \vdots & & & \vdots \\ \vdots & \vdots & & & \vdots \\ * & * & \cdots & \cdots & * \end{bmatrix}$$

$$
\xrightarrow{G_{23}G_{13}G_{12}A}
\begin{bmatrix}
* & * & * & \cdots & * \\
0 & * & * & & * \\
0 & 0 & * & & * \\
* & * & * & & * \\
\vdots & & & & \vdots \\
\vdots & & & & \vdots \\
\vdots & & & & \vdots \\
\vdots & & & & \vdots \\
* & * & * & \cdots & *
\end{bmatrix}
\cdots\longrightarrow
\begin{bmatrix}
* & \cdots & \cdots & \cdots & * \\
0 & \ddots & & & \vdots \\
0 & \ddots & \ddots & & \vdots \\
\vdots & & \ddots & \ddots & \vdots \\
\vdots & & & \ddots & * \\
0 & \cdots & \cdots & \cdots & 0 \\
* & \cdots & \cdots & \cdots & * \\
\vdots & & & & \vdots \\
* & \cdots & \cdots & \cdots & *
\end{bmatrix}
$$

$$
\cdots\longrightarrow
\begin{bmatrix}
* & \cdots & \cdots & \cdots & * \\
0 & \ddots & & & \vdots \\
\vdots & \ddots & \ddots & & \vdots \\
\vdots & & \ddots & \ddots & \vdots \\
\vdots & & & \ddots & * \\
\vdots & & & & 0 \\
\vdots & & & & \vdots \\
\vdots & & & & \vdots \\
0 & \cdots & \cdots & \cdots & 0
\end{bmatrix}.
$$

至此可在$2n^2(m-n/3)$次运算内完成.

详细的算法描述可参见[GC1989]的 146-157 页. □

计算矩阵$A\in\mathbb{R}^{n\times n}$特征值问题的非常重要的 QR 方法,基本迭代格式是

$$A_k=Q_kR_k,\quad A_{k+1}=R_kQ_k. \tag{7.14}$$

即每一次迭代的主要运算量是 QR 分解 Q_kR_k,然后计算 R_kQ_k.如果直接从 A 出发,从上述定理知道,每一次迭代即对每个 k,(7.14)的运算量是 $O(n^3)$.

但是,在实际计算中,常依 **5.6.2**,应用 Householder 变换,首先将 A 约化为与其相似的 Hessenberg 标准形 H,运算量为 $O(n^3)$;然后从 H 出发进行 QR 迭代,QR 分解用 Givens 变换实现,即相继取 $n-1$ 个 Givens 变换

$$G_i \equiv G(i,i+1,\theta_i), \quad i=1,\cdots,n-1, \tag{7.15}$$

使得

$$G_{n-1}\cdots G_2G_1H_k = R_k \tag{7.16}$$

为上三角矩阵,再计算

$$H_{k+1} = R_kG_1^{\mathrm{T}}G_2^{\mathrm{T}}\cdots G_{n-1}^{\mathrm{T}}, \tag{7.17}$$

每一次迭代,(7.16)和(7.17)在一起的运算量为 $4n^2$.此时,

$$Q_k = G_1^{\mathrm{T}}G_2^{\mathrm{T}}\cdots G_{n-1}^{\mathrm{T}}, \tag{7.18}$$

运算量为 $O(n)$.这样,使 QR 方法的运算量大为减少.

5.8 Gauss 变换和 LU 分解

5.8.1 定义 设 $x=(x_1,\cdots,x_n)^{\mathrm{T}}\in\mathbb{R}^n$, $x_k\neq 0$.记

$$l_{ik} = x_i/x_k, \quad i=k+1,\cdots,n, \tag{8.1}$$

向量

$$l^{(k)} = (\underbrace{0,\cdots,0}_{k},l_{k+1,k},\cdots,l_{nk}) \tag{8.2}$$

称为相伴 x 的 **Gauss 向量**.矩阵

$$L_k = I - l^{(k)} e_k^{\mathrm{T}} = \begin{bmatrix} 1 & & & & & \\ & \ddots & & & 0 & \\ & & 1 & & & \\ & & -l_{k+1,k} & 1 & & \\ & & \vdots & & \ddots & \\ 0 & & -l_{nk} & 0 & & 1 \end{bmatrix} \tag{8.3}$$

称为相伴 x 的 **Gauss 变换**,也称**初等下三角矩阵**.

Householder 变换和 GIvens 变换都是正交变换,它们都是约化向量的某些元素为零的基本工具.Gauss 变换也是这样的基本工具,但它不是正交变换.

5.8.2 引理 设 L_k 是相伴 $x=(x_1,\cdots,x_n)^{\mathrm{T}}$ 的 Gauss 变换, 则

$$L_k x = (x_1,\cdots,x_k,\underbrace{0,\cdots,0}_{n-k})^{\mathrm{T}}. \tag{8.4}$$

证 依 **5.8.1**,

$$L_k x = Ix - l^{(k)} e_k^T x = x - x_k l^{(k)} = (x_1,\cdots,x_k,0,\cdots,0)^{\mathrm{T}}. \quad \square$$

5.8.3 引理 设 $L_1,\cdots,L_{n-1}$ 是相伴 $x=(x_1,\cdots,x_n)^{\mathrm{T}}$ 的 Gauss 变换, 则

$$L_k^{-1} = I + l^{(k)} e_k^{\mathrm{T}} = \begin{bmatrix} 1 & & & & & \\ & \ddots & & & 0 & \\ & & 1 & & & \\ & & l_{k+1,k} & 1 & & \\ & & \vdots & & \ddots & \\ 0 & & l_{nk} & 0 & & 1 \end{bmatrix},$$

$$k=1,\cdots,n-1. \tag{8.5}$$

而且

$$L \equiv L_1^{-1} \cdots L_{n-1}^{-1} = \begin{bmatrix} 1 & & & & \\ l_{21} & 1 & & \text{\Huge 0} & \\ l_{31} & l_{32} & \ddots & & \\ \vdots & \vdots & & 1 & \\ l_{n1} & l_{n2} & \cdots & l_{n,n-1} & 1 \end{bmatrix}. \tag{8.6}$$

证 注意到 $e_j^{\mathrm{T}} l^{(k)} = 0, j \le k$,有

$$\begin{aligned}\left(I - l^{(k)} e_k^{\mathrm{T}}\right)\left(I + l^{(k)} e_k^{\mathrm{T}}\right) &= I - \left(l^{(k)} e_k^{\mathrm{T}}\right)\left(l^{(k)} e_k^{\mathrm{T}}\right) \\ &= I - l^{(k)}\left(e_k^{\mathrm{T}} l^{(k)}\right) e_k^{\mathrm{T}} = I - \left(e_k^{\mathrm{T}} l^{(k)}\right) l^{(k)} e_k^{\mathrm{T}} = I,\end{aligned}$$

这就证明了(8.5).而且,因此有

$$\begin{aligned}L &\equiv L_1^{-1} \cdots L_{n-1}^{-1} = \left(I + l^{(1)} e_1^{\mathrm{T}}\right) \cdots \left(I + l^{(n-1)} e_{n-1}^{\mathrm{T}}\right) \\ &= I + l^{(1)} e_1^{\mathrm{T}} + \cdots + l^{(n-1)} e_{n-1}^{\mathrm{T}},\end{aligned}$$

此即(8.6). □

5.8.4 LU 分解定理 设 $A = [a_{ij}] \in \mathbb{R}^{m \times n}, s = \min\{m-1, n\}$, A 的主子矩阵

$$A_k \equiv \begin{bmatrix} a_{11} & \cdots & a_{1k} \\ \cdots & \cdots & \cdots \\ a_{k1} & \cdots & a_{kk} \end{bmatrix}, \quad k = 1, \cdots, s \tag{8.7}$$

均非奇异,则存在单位下三角矩阵

$$L = \begin{bmatrix} 1 & & \text{\Huge 0} & \\ l_{21} & \ddots & & \\ \vdots & \ddots & \ddots & \\ l_{m1} & \cdots & l_{m,m-1} & 1 \end{bmatrix} \in \mathbb{R}^{m \times m} \tag{8.8}$$

和上三角矩阵 $U \in \mathbb{R}^{m \times n}$,

$$U=\begin{bmatrix}U_1\\0\end{bmatrix},\text{当}\, m\ge n;\ U=\begin{bmatrix}U_1 & 0\end{bmatrix},\text{当}\, m<n,\tag{8.9}$$

其中

$$U_1=\begin{bmatrix}u_{11} & u_{12} & \cdots & u_{1r}\\ & u_{22} & \cdots & u_{2r}\\ & & \ddots & \vdots\\ 0 & & & u_{rr}\end{bmatrix},\ r=\min\{m,n\},\tag{8.10}$$

使得

$$A=LU,\tag{8.11}$$

而且

$$\det A_k=u_{11}\cdots u_{kk},\ k=1,\cdots,s.\tag{8.12}$$

证 由于 $a_{11}=\det A_1\neq 0$,可以确定 Gauss 变换

$$L_1=I-l^{(1)}e_1^T=\begin{bmatrix}1 & & 0 & \\ -l_{21} & 1 & & \\ \vdots & & \ddots & \\ -l_{m1} & 0 & & 1\end{bmatrix}\in\mathbb{R}^{m\times m},$$

其中

$$l_{i1}=a_{i1}/a_{11},\ i=2,\cdots,m.$$

依 **5.8.2**,

$$A^{(1)}\equiv L_1A=\begin{bmatrix}A_{11}^{(1)} & A_{12}^{(1)}\\ 0 & A_{22}^{(1)}\end{bmatrix},$$

其中 $A_{11}^{(1)}=[a_{11}]$,而 $A_{12}^{(1)}\in\mathbb{R}^{1\times(n-1)}$,

$$A_{22}^{(1)}=\begin{bmatrix}a_{22}^{(1)} & \cdots & a_{2n}^{(1)}\\ \cdots & \cdots & \cdots\\ a_{m2}^{(1)} & \cdots & a_{mn}^{(1)}\end{bmatrix},$$

容易推出 $\det A_2 = a_{11}a_{22}^{(1)}$,由假设 $\det A_2 \neq 0$,故 $a_{22}^{(1)} \neq 0$.因此,仍可用 Gauss 变换,将 $A^{(1)}$的第2列($A_{22}^{(1)}$的第1列)中的 $a_{22}^{(1)}$ 以下的元素约化为零.

现假设已确定 Gauss 变换 $L_1,\cdots,L_{k-1} \in \mathbb{R}^{m\times m}$,使得

$$A^{(k-1)} \equiv L_{k-1}\cdots L_1 A = \begin{bmatrix} A_{11}^{(k-1)} & A_{12}^{(k-1)} \\ 0 & A_{22}^{(k-1)} \end{bmatrix}, \qquad (8.13)$$

其中 $A_{11}^{(k-1)} \in \mathbb{R}^{(k-1)\times(k-1)}$ 是上三角矩阵, $A_{12}^{(k-1)} \in \mathbb{R}^{(k-1)\times(n-k+1)}$,

$$A_{22}^{(k-1)} = \begin{bmatrix} a_{kk}^{(k-1)} & \cdots & a_{kn}^{(k-1)} \\ \cdots & \cdots & \cdots \\ a_{mk}^{(k-1)} & \cdots & a_{mn}^{(k-1)} \end{bmatrix}, \qquad (8.14)$$

并假定 $a_{kk}^{(k-1)} \neq 0$.

那么,可以确定 Gauss 变换

$$L_k = I - l^{(k)}e_k^T = \begin{bmatrix} 1 & & & & & \\ & \ddots & & & 0 & \\ & & 1 & & & \\ & & -l_{k+1,k} & 1 & & \\ & & \vdots & & \ddots & \\ 0 & & -l_{mk} & 0 & & 1 \end{bmatrix} \in \mathbb{R}^{m\times m}, \qquad (8.15)$$

其中

$$l_{ik} = a_{ik}^{(k-1)}/a_{kk}^{(k-1)}, \quad i = k+1,\cdots,m. \qquad (8.16)$$

于是

$$A^{(k)} \equiv L_k A^{(k-1)} = L_k\cdots L_1 A = \begin{bmatrix} A_{11}^{(k)} & A_{12}^{(k)} \\ 0 & A_{22}^{(k)} \end{bmatrix}, \qquad (8.17)$$

这里 $A_{11}^{(k)} \in \mathbb{R}^{k\times k}$ 是上三角矩阵.由于

$$A = L_1^{-1} \cdots L_k^{-1} A^{(k)} = \left(L_k \cdots L_1\right)^{-1} A^{(k)},$$

注意到有

$$A^{(k)} = \begin{bmatrix} I_k & 0 \\ 0 & A_{22}^{(k)} \end{bmatrix} \begin{bmatrix} A_{11}^{(k)} & A_{12}^{(k)} \\ 0 & I_{n-k} \end{bmatrix}$$

和

$$\left(L_k \cdots L_1\right)^{-1} = L_1^{-1} \cdots L_k^{-1} = \prod_{i=1}^{k} \left(I_m + l^{(i)} e_i^{\mathrm{T}}\right)$$

$$= I_m + \begin{bmatrix} l^{(1)} & \cdots & l^{(k)} & 0 & \cdots & 0 \end{bmatrix} \equiv \begin{bmatrix} L_{11}^{(k)} & 0 \\ L_{21}^{(k)} & I_{m-k} \end{bmatrix},$$

因而

$$A = \begin{bmatrix} L_{11}^{(k)} & 0 \\ L_{21}^{(k)} & I_{m-k} \end{bmatrix} \begin{bmatrix} I_k & 0 \\ 0 & A_{22}^{(k)} \end{bmatrix} \begin{bmatrix} A_{11}^{(k)} & A_{12}^{(k)} \\ 0 & I_{n-k} \end{bmatrix},$$

由此,得

$$A_k = L_{11}^{(k)} A_{11}^{(k)}.$$

因为

$$\det A_{11}^{(k)} = a_{kk}^{(k-1)} \det A_{11}^{(k-1)},$$

推出

$$\det A_k = \det L_{11}^{(k)} \det A_{11}^{(k)} = \det A_{11}^{(k)} = a_{11}^{(0)} \cdots a_{kk}^{(k-1)},$$

这里规定 $a_{11}^{(0)} \equiv a_{11}$.前主子矩阵(8.7)非奇异的假设蕴涵约化过程可以继续到 $k = s$.其时,

$$U \equiv A^{(s)} \in \mathbb{R}^{m \times n}$$

是上三角矩阵,而

$$L \equiv L_1^{-1} \cdots L_s^{-1} \in \mathbb{R}^{m \times m}$$

是单位下三角矩阵,并且 $A = LU$. □

(8.11)称为矩阵 A 的 **LU 分解**.

一般,如果 $A \in \mathbb{C}^{n\times n}$ 的所有前主子矩阵是非奇异的,则称 A 为**强非奇异矩阵**(strongly nonsingular matrix).

现在,假定 $A \in \mathbb{R}^{n\times n}$ 是给定的非奇异矩阵.考虑求解线性方程组

$$Ax = b, \tag{8.18}$$

其中 $b \in \mathbb{R}^n$ 是给定向量, $x \in \mathbb{R}^n$ 是待求向量.如果 A 是强非奇异的,那么依 **5.8.4**, A 存在 LU 分解 $A = LU$.于是

$$Ax = (LU)x = L(Ux) = b,$$

由此,若令 $Ux = y$,则 $Ly = b$.这样,求解方程组归结为先求解下三角形方程组

$$Ly = b, \tag{8.19}$$

称为**向前消去**,然后求解上三角形方程组

$$Ux = y, \tag{8.20}$$

称为**回代**.求解(8.19)和(8.20)都是简单的递推过程.这就是利用矩阵分解语言来表述的 Gauss 消去法.

6 特型矩阵

6.1 带状矩阵

6.1.1 引言 “特型矩阵(patterned matrix)”泛指元素呈现可识别模式的矩阵.

这样的矩阵是大量存在的.它们有不少分布在本书的其它章节之中.例如,对称矩阵、自伴矩阵，以及相应的斜对称矩阵和斜自伴矩阵;Jordan 形、Hessenberg 形;还有友矩阵、Vandermonde 矩阵,等等.自然,在所有特殊形式的矩阵中,最简单的莫过于对角矩阵、双对角矩阵和三对角矩阵,进而是上和下三角矩阵,它们都归属于带状矩阵.

本章集中一批从视觉上可识别其模式的矩阵.“从视觉上可识别”的含义是不言而喻的,举例来说,自伴矩阵一眼便可看出;正定矩阵是自伴矩阵的子类,但一般却不可能直接看出一个自伴矩阵是不是正定的.

关注特型矩阵有两个理由:

第一,每种特型矩阵比起任意矩阵几乎总是至少有一个特殊性质更容易加以确定.例如,对角或三角矩阵的特征值是明摆着的;友矩阵的特征多项式可随手写出;而 Vandermonde 矩阵的行列式有简单的显式表示;许多特型矩阵的逆矩阵的计算比一般矩阵要省力些;如此等等.

第二,所研究的特型矩阵是自然地出自广泛的各种应用之中的,包括控制与系统理论、信号处理、统计以及微分方程数值解等.

6.1.2 定义 设 $A=[a_{ij}]\in\mathbb{C}^{n\times n}$ 所有非零元素包含在由主对角线及其之上的 p 条对角线与其之下的 q 条对角线组成的**带**内.利用元素来描述,即有

$$a_{ij}=0,\ \forall i,j=1,\cdots,n,\quad j-i>p \text{ 或 } i-j>q. \tag{1.1}$$

具体地, A 形如

$$\begin{bmatrix} * & \cdots & * & & & 0 \\ \vdots & \ddots & & \ddots & & \\ \vdots & & \ddots & & \ddots & \\ * & & & \ddots & & * \\ & \ddots & & & \ddots & \vdots \\ 0 & & * & \cdots & \cdots & * \end{bmatrix} \begin{matrix} \\ \\ \\ \left.\begin{matrix} \\ \\ \\ \end{matrix}\right\} p+1 \end{matrix}$$

$$\underbrace{\qquad\qquad\qquad}_{q+1}$$

则称 A 为 (p,q)-**带状矩阵**(band matrix).

常会遇到带状矩阵,在常或偏微分方程数值求解中尤其如此.

例 考虑两点边值问题

$$\begin{cases} y''(x)+q(x)y(x)=f(x), & 0<x<1, \\ y(0)=\alpha, & y(1)=\beta, \end{cases} \tag{1.2}$$

其中 q 和 f 是已知实函数, $\alpha,\beta\in\mathbb{R}$.将问题离散化,把区间 $[0,1]$ 分成 $n+1$ 等份,令

$$h=1/(n+1);\quad x_i=ih,\ i=0,1,\cdots,n+1.$$

利用差商逼近二阶导数:

$$y''(x_i)\approx\frac{y_{i+1}-2y_i+y_{i-1}}{h^2},\ i=1,\cdots,n.$$

问题(1.2)近似地替换成如下线性方程组

$$Ay=h^2f+b, \tag{1.3}$$

其中系数矩阵

$$A=\begin{bmatrix} 2+h^2q_1 & -1 & & 0 \\ -1 & \ddots & \ddots & \\ & \ddots & \ddots & -1 \\ 0 & & -1 & 2+h^2q_n \end{bmatrix} \tag{1.4}$$

是$(1,1)$-带状矩阵,

$$q_i=q(x_i),\ i=1,\cdots,n.$$

向量

$$f=(f_1,\cdots,f_n)^{\mathrm{T}};\ f_i=f(x_i),\ i=1,\cdots,n$$

和

$$b=(y_0,0,\cdots,0,y_{n+1})^{\mathrm{T}},$$

其中

$$y_0=y(x_0)=\alpha,\ y_{n+1}=y(x_{n+1})=\beta,$$

而

$$y=(y_1,\cdots,y_n)^{\mathrm{T}};\ y_i\approx y(x_i),\ i=1,\cdots,n$$

是待求的问题(1.2)的数值解.

常见的带状矩阵(或对应的块带状矩阵)有:

$(0,0)$-带状矩阵,就是**对角矩阵**;

$(1,0)$-带状矩阵和$(0,1)$-带状矩阵,分别称为**上双对角矩阵**和**下双对角矩阵**;

$(1,1)$-带状矩阵,就是**三对角矩阵**;

$(2,2)$-带状矩阵,称为**五对角矩阵**;

$n\times n$的$(n-1,0)$-带状矩阵和$(0,n-1)$-带状矩阵,分别就是**上三角矩阵**和**下三角矩阵**;

$n\times n$的$(n-1,1)$-带状矩阵和$(1,n-1)$-带状矩阵,分别就是**上Hessenberg矩阵**和**下Hessenberg矩阵**.

下面着重讨论三对角矩阵.从元素看,$A=[a_{ij}]\in\mathbb{C}^{n\times n}$满足

$$a_{ij}=0;\quad \forall i,j=1,\cdots,n,\ |i-j|>1 \tag{1.5}$$

时是三对角矩阵.为了使用方便,n 阶三对角矩阵常用如下表示:

$$A_n=\begin{bmatrix} a_1 & b_1 & & \text{\Large 0} \\ c_1 & a_2 & \ddots & \\ \text{\Large 0} & \ddots & \ddots & b_{n-1} \\ & & c_{n-1} & a_n \end{bmatrix}. \tag{1.6}$$

6.1.3 定理 形如(1.6)的三对角矩阵 A_n 的行列式可按如下递推公式求值:

$$\det A_k=a_k\det A_{k-1}-b_{k-1}c_{k-1}\det A_{k-2},\ k=1,2,\cdots,n. \tag{1.7}$$

这里规定 $\det A_{-1}=0,\det A_0=1$.

容易利用数学归纳法和行列式的 Laplace 展开 **1.5.11** 给出定理的证明,留作练习.

这一定理表明三对角矩阵的行列式计算比较简单.

6.1.4 定义 矩阵

$$A=\begin{bmatrix} 1 & 1 & & \text{\Large 0} \\ -1 & \ddots & \ddots & \\ \text{\Large 0} & \ddots & \ddots & 1 \\ & & -1 & 1 \end{bmatrix}\in\mathbb{R}^{n\times n} \tag{1.8}$$

称为 **Fibonacci 矩阵**.

利用(1.7)即知(1.8)中的 A 的行列式正好是 **Fibonacci 序列**

$$\alpha_1=1,\alpha_2=2;$$

$$\alpha_k=\alpha_{k-1}+\alpha_{k-2},\ k=3,4,\cdots \tag{1.9}$$

的第 n 项. Fibonacci 序列 $\{\alpha_k\}$ 的头几项是:

$$1,\quad 2,\quad 3,\quad 5,\quad 8,\quad 13,\quad \cdots$$

6.1.5 定理 设 $A \equiv A_n \in \mathbb{R}^{n\times n}$ 是形如(1.6)的三对角矩阵,其元素具有正负号对称性质:

$$b_i c_i > 0,\ i = 1,\cdots,n-1, \tag{1.10}$$

则 A 相似于实对称矩阵, A 的特征值全为实数.

而且,在条件

$$b_i c_i \geq 0,\ i = 1,\cdots,n-1 \tag{1.11}$$

下, A 的特征值全为实数的结论仍然成立.

证 实对称矩阵的所有特征值必为实数,因此只须证明在条件(1.10)之下, A 相似于实对称矩阵.考虑用对角矩阵对 A 作相似变换.设 $D = \operatorname{diag}(d_1,\cdots,d_n)$ 是非奇异的,则

$$DAD^{-1} = \begin{bmatrix} a_1 & \frac{d_1}{d_2}b_1 & & & \\ \frac{d_2}{d_1}c_1 & a_2 & \frac{d_2}{d_3}b_2 & & \\ & \frac{d_3}{d_2}c_2 & \ddots & \ddots & \\ & & \ddots & \ddots & \frac{d_{n-1}}{d_n}b_{n-1} \\ & & & \frac{d_n}{d_{n-1}}c_{n-1} & a_n \end{bmatrix}.$$

欲使其成为对称的,亦即

$$\frac{d_{i-1}}{d_i}b_{i-1} = \frac{d_i}{d_{i-1}}c_{i-1},\ i = 2,\cdots,n.$$

如此关系式,在(1.10)及附加

$$d_i > 0,\ i = 1,\cdots,n$$

的条件下,等价于

$$d_i = \sqrt{\frac{b_{i-1}}{c_{i-1}}} d_{i-1}, \ i = 2, \cdots, n, \tag{1.12}$$

这是关于产生 D 的对角元素的递推公式,其中 d_1 可以是任一正数,比如不妨取 $d_1 = 1$.因此,使 DAD^{-1} 实对称化的实对角矩阵 D 是存在的.

现在将条件从(1.10)减弱为(1.11).任意取定一个 $\varepsilon > 0$.对于 $b_i c_i = 0$ 的情形:

(1) 如果 $b_i = c_i = 0$,则可将 A 分成独立的低阶矩阵;

(2) 如果 $b_i \neq 0, c_i = 0$,则用 $(\operatorname{sgn} b_i)\varepsilon$ 来代替 c_i;

(3) 如果 $b_i = 0, c_i \neq 0$,则用 $(\operatorname{sgn} c_i)\varepsilon$ 来代替 b_i.

作如此处理后的矩阵,或是单个符合条件(1.10)的矩阵,或是符合条件(1.10)的若干低阶矩阵的直和.

根据定理前一半结论的证明,得知处理后矩阵的特征值全为实数.因此,A 作为处理后矩阵当 $\varepsilon \to 0$ 时的极限情形,其特征值也必全为实数. □

6.2 轮换矩阵

6.2.1 定义 设 $c_0, c_1, \cdots, c_{n-1} \in \mathbb{C}$, $n \times n$ 矩阵

$$C = \begin{bmatrix} c_0 & c_1 & c_2 & \cdots & c_{n-1} \\ c_{n-1} & c_0 & c_1 & \cdots & c_{n-2} \\ c_{n-2} & c_{n-1} & c_0 & \cdots & c_{n-3} \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ c_1 & c_2 & c_3 & \cdots & c_0 \end{bmatrix} \tag{2.1}$$

称为数组 $c_0,c_1,\cdots,c_{n-1}$ 的**轮换矩阵**(circulant matrix).

轮换矩阵(2.1)中的每一行是由前一行的元素右移一个位置,并将“溢出”元素挪到左边第1个位置而组成的.

轮换矩阵 C 完全由第1行来确定,在有的著作(如[D1979])里使用记号

$$\operatorname{circ}(c_0,c_1,\cdots,c_{n-1}). \tag{2.2}$$

从(2.1)直接看出,轮换矩阵 C 沿平行主对角线的每一对角线上的元素是相等的,C 是关于次对角线对称的.

6.2.2 定义 置换矩阵

$$P=\begin{bmatrix} 0 & 1 & 0 & \cdots & 0 \\ \vdots & 0 & 1 & \ddots & \vdots \\ \vdots & & \ddots & \ddots & 0 \\ 0 & & & \ddots & 1 \\ 1 & 0 & \cdots & \cdots & 0 \end{bmatrix} \in \mathbb{R}^{n\times n} \tag{2.3}$$

是重要的特殊轮换矩阵,称为**循环置换矩阵**(cyclic permutation matrix)或**基本轮换矩阵**或**移位矩阵**(shift matrix).

利用(2.2)的记号,

$$P=\operatorname{cire}(0,1,0,\cdots,0).$$

6.2.3 定理 设 $P\in\mathbb{R}^{n\times n}$ 是基本轮换矩阵,则

(1) $P^n=I$.

(2) P 的特征值为

$$\mu_k=\mathrm{e}^{\mathrm{i}\frac{2\pi(k-1)}{n}},\quad k=1,\cdots n. \tag{2.4}$$

(3) P 与任一轮换矩阵 $C\in\mathbb{C}^{n\times n}$ 可交换:

$$CP = PC. \tag{2.5}$$

(4) $C \in \mathbb{C}^{n\times n}$ 为轮换矩阵的充分必要条件是 C 可以表示为

$$C = \sum_{k=0}^{n-1} c_k P^k, \tag{2.6}$$

这里规定 $P^0 = I$.

证 (1)和(3)可直接验证.

(2) P 的特征多项式是 $\lambda^n - 1$,它有 n 个不同的根,即 P 的特征值:

$$1,\quad \theta,\quad \theta^2,\quad \cdots,\quad \theta^{n-1};\qquad \theta = \exp\left(\mathrm{i}\frac{2\pi}{n}\right),$$

(4) 直接从定义 **6.2.1** 和 **6.2.2** 推出. □

6.2.4 定理 设 $B, C \in \mathbb{C}^{n\times n}$ 是轮换矩阵,则

(1) BC 是轮换矩阵.

(2) B 和 C 可交换: $BC = CB$.

证 (1) 从 **6.2.3** 的(1)和(4)推出.

(2) 从 **6.2.3** 的(3)和(4)推出. □

6.2.5 定理 设 $C = \mathrm{circ}(c_0, c_1, \cdots, c_{n-1})$,则

(1) C 的特征值是

$$\lambda_k = \phi(\mu_k),\ \ k = 1, \cdots, n. \tag{2.7}$$

而且属于 λ_k 的特征向量可取为

$$\left(1, \mu_k, \mu_k^2, \cdots, \mu_k^{n-1}\right)^{\mathrm{T}},\ \ k = 1, \cdots, n. \tag{2.8}$$

(2) C 的行列式

$$\det C = \prod_{k=1}^{n} \phi(\mu_k), \tag{2.9}$$

其中

$$\phi(t) = c_0 + c_1 t + c_2 t^2 + \cdots + c_{n-1} t^{n-1} \tag{2.10}$$

而 $\mu_1,\cdots,\mu_n$ 由(2.4)给出.

证 (1) 由(2.4)和(2.6)推出(2.7).

而且,容易验证成立

$$CV = V\mathrm{diag}(\phi(\mu_1),\cdots,\phi(\mu_n)), \tag{2.11}$$

其中 $V \in \mathbb{C}^{n\times n}$ 是关于 $\mu_1,\cdots,\mu_n$ 的 Vandermonde 矩阵:

$$V = \begin{bmatrix} 1 & 1 & 1 & \cdots & 1 \\ \mu_1 & \mu_2 & \mu_3 & \cdots & \mu_n \\ \mu_1^2 & \mu_2^2 & \mu_3^2 & \cdots & \mu_n^2 \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ \mu_1^{n-1} & \mu_2^{n-1} & \mu_3^{n-1} & \cdots & \mu_n^{n-1} \end{bmatrix}. \tag{2.12}$$

(2) $\det C = \prod_{k=1}^{n} \lambda_k$ 并利用(2.7). □

值得指出的是:V 和 $c_0,c_1,\cdots,c_{n-1}$ 无关.这表明所有轮换矩阵都可通过 V 来对角化.

换言之,所有轮换矩阵有共同的线性无关的(可取为正交的)特征向量组.

6.2.6 定义 设 $c_0,c_1,\cdots,c_{n-1},\alpha \in \mathbb{C}$, $n\times n$ 矩阵

$$\tilde{C} = \begin{bmatrix} c_0 & c_1 & c_2 & \cdots & c_{n-1} \\ \alpha c_{n-1} & c_0 & c_1 & & c_{n-2} \\ \alpha c_{n-2} & \alpha c_{n-1} & c_0 & \ddots & \vdots \\ \vdots & & \ddots & \ddots & c_1 \\ \alpha c_1 & \alpha c_2 & \cdots & \alpha c_{n-1} & c_0 \end{bmatrix} \tag{2.13}$$

称为**广义轮换矩阵**.特别,当 $\alpha = -1$ 时称为**斜轮换矩阵**.

6.2.7 定理 设 $\tilde{C} \in \mathbb{C}^{n\times n}$ 是形如(2.13)的广义轮换矩阵,则

(1) $\tilde{C}$ 的特征值是

$$\lambda_k = \phi(\tilde{\mu}_k),\ k = 1,\cdots,n. \tag{2.14}$$

(2) $\tilde{C}$ 的行列式

$$\det \tilde{C} = \prod_{k=1}^{n} \phi(\tilde{\mu}_k), \tag{2.15}$$

其中 ϕ 仍由(2.10)给出,而

$$\tilde{\mu}_k = \alpha^{\frac{1}{n}} e^{i\frac{2\pi(k-1)}{n}},\ k = 1,\cdots,n. \tag{2.16}$$

证 与(2.6)相仿,成立

$$\tilde{C} = \sum_{k=0}^{n-1} c_k \tilde{P}^k, \tag{2.17}$$

其中

$$\tilde{P} = \begin{bmatrix} 0 & 1 & 0 & \cdots & 0 \\ \vdots & 0 & 1 & \ddots & \vdots \\ \vdots & & \ddots & \ddots & 0 \\ 0 & & & \ddots & 1 \\ \alpha & 0 & \cdots & \cdots & 0 \end{bmatrix} \in \mathbb{C}^{n\times n}, \tag{2.18}$$

$\tilde{P}$ 的特征多项式是

$$\lambda^n - \alpha,$$

推出其特征值为(2.16).

因此,从(2.17)即得(2.14)及(2.15). □

6.3 Toeplitz 矩阵

6.3.1 定义 $T = [t_{ij}] \in \mathbb{C}^{n\times n}$ 称为 **Toeplitz 矩阵**,如果满足

$$t_{ij}=t_{j-i},\ i,j=1,\cdots,n. \tag{3.1}$$

具体地,形如

$$T=\begin{bmatrix} t_0 & t_1 & t_2 & \cdots & t_{n-2} & t_{n-1} \\ t_{-1} & t_0 & t_1 & \ddots & & t_{n-2} \\ t_{-2} & t_{-1} & t_0 & \ddots & \ddots & \vdots \\ \vdots & \ddots & \ddots & \ddots & \ddots & t_2 \\ t_{-n+2} & & \ddots & \ddots & \ddots & t_1 \\ t_{-n+1} & t_{-n+2} & \cdots & t_{-2} & t_{-1} & t_0 \end{bmatrix}. \tag{3.2}$$

T 完全由其第1行和第1列的 $2n-1$ 个元素

$$t_{n-1},\cdots,t_1,t_0,t_{-1},\cdots,t_{-n+1}\in\mathbb{C}$$

确定,也说是由**拟多项式**

$$p(\lambda)\equiv t_{n-1}\lambda^{n-1}+\cdots+t_1\lambda+t_0+t_{-1}\lambda^{-1}+\cdots+t_{-n+1}\lambda^{-n+1} \tag{3.3}$$

生成的.

Toeplitz矩阵沿平行主对角线的每一对角线上的元素是相等的,是关于次对角线对称的.

特别,轮换矩阵是 Toeplitz 矩阵.

简单的 Toeplitz 矩阵

$$F=\begin{bmatrix} 0 & 1 & & \text{\huge 0} \\ & \ddots & \ddots & \\ \text{\huge 0} & & \ddots & 1 \\ & & & 0 \end{bmatrix}\in\mathbb{R}^{n\times n} \tag{3.4}$$

和

$$B=\begin{bmatrix}0 & & & 0\\ 1 & \ddots & & \\ & \ddots & \ddots & \\ 0 & & 1 & 0\end{bmatrix}\in\mathbb{R}^{n\times n}, \qquad (3.5)$$

因它们作用于标准基$\{e_1,\cdots,e_n\}$时所产生的直观影响,分别称为**前向移位矩阵**(forward shift matrix)和**后向移位矩阵**(backward shift matrix).

Toeplitz 矩阵自然地出现在微分方程数值解、二角矩量等诸问题中.

6.3.2 定义 $A=[a_{ij}]\in\mathbb{C}^{n\times n}$称为**广对称矩阵**(persymmetric matrix),如果A是关于次对角线对称的,即满足

$$a_{ij}=a_{n+1-j,n+1-i},\quad i,j=1,\cdots,n. \qquad (3.6)$$

等价于

$$A=EA^{\mathrm{T}}E, \qquad (3.7)$$

其中

$$E=[e_n \quad \cdots \quad e_1]$$

是反序单位矩阵.

6.3.3 定理 $T\in\mathbb{C}^{n\times n}$为 Toeplitz 矩阵的充分必要条件是$T$可以表示为

$$T=\sum_{k=1}^{n-1}t_{-k}B^k+\sum_{k=0}^{n-1}t_kF^k, \qquad (3.8)$$

其中F和B由(3.4)和(3.5)给定.

证明非常容易,留作练习.

6.3.4 定理 非奇异矩阵$T\in\mathbb{C}^{n\times n}$为 Toeplitz 矩阵的充分必要条件是存在$n\times n$友矩阵

$$C^{(1)}=\begin{bmatrix} p_{n-1} & p_{n-2} & \cdots & p_1 & p_0 \\ 1 & 0 & \cdots & \cdots & 0 \\ 0 & \ddots & \ddots & & \vdots \\ \vdots & \ddots & \ddots & \ddots & \vdots \\ 0 & \cdots & 0 & 1 & 0 \end{bmatrix} \tag{3.9}$$

和

$$C^{(2)}=\begin{bmatrix} 0 & \cdots & \cdots & 0 & p_0 \\ 1 & \ddots & & \vdots & p_1 \\ 0 & \ddots & \ddots & \vdots & \vdots \\ \vdots & \ddots & \ddots & 0 & p_{n-2} \\ 0 & \cdots & 0 & 1 & p_{n-1} \end{bmatrix}, \tag{3.10}$$

使得

$$C^{(1)}T = TC^{(2)}. \tag{3.11}$$

证 设$T=[t_{ij}]$.

直接计算得

$$C^{(1)}T=\begin{bmatrix} \sum_{i=1}^{n} t_{i1}p_{n-i} & \sum_{i=1}^{n} t_{i2}p_{n-i} & \cdots & \sum_{i=1}^{n} t_{i,n-1}p_{n-i} & \sum_{i=1}^{n} t_{in}p_{n-i} \\ t_{11} & t_{12} & \cdots & t_{1,n-1} & t_{1n} \\ t_{21} & t_{22} & \cdots & t_{2,n-1} & t_{2n} \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ t_{n-1,1} & t_{n-1,2} & \cdots & t_{n-1,n-1} & t_{n-1,n} \end{bmatrix} \tag{3.12}$$

和

$$TC^{(2)}=\begin{bmatrix} t_{12} & t_{13} & \cdots & t_{1n} & \sum_{j=1}^{n} t_{1j}p_{j-1} \\ t_{22} & t_{23} & \cdots & t_{2n} & \sum_{j=1}^{n} t_{2j}p_{j-1} \\ t_{32} & t_{33} & \cdots & t_{3n} & \sum_{j=1}^{n} t_{3j}p_{j-1} \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ t_{n2} & t_{n3} & \cdots & t_{nn} & \sum_{j=1}^{n} t_{nj}p_{j-1} \end{bmatrix}. \tag{3.13}$$

如果T是非奇异 Toeplitz 矩阵,则成立(3.1),T可表示为

$$T=\begin{bmatrix} t_0 & t_1 & t_2 & \cdots & t_{n-2} & t_{n-1} \\ t_{-1} & t_0 & t_1 & \ddots & & t_{n-2} \\ t_{-2} & t_{-1} & t_0 & \ddots & \ddots & \vdots \\ \vdots & \ddots & \ddots & \ddots & \ddots & t_2 \\ t_{-n+2} & & \ddots & \ddots & \ddots & t_1 \\ t_{-n+1} & t_{-n+2} & \cdots & t_{-2} & t_{-1} & t_0 \end{bmatrix}.$$

而且根据(3.7),

$$T=ET^{\mathrm{T}}E, \tag{3.14}$$

(3.12)和(3.13)成为

$$C^{(1)}T$$

$$=\begin{bmatrix} \sum_{i=1}^{n} t_{1-i}p_{n-i} & \sum_{i=1}^{n} t_{2-i}p_{n-i} & \cdots & \sum_{i=1}^{n} t_{n-1-i}p_{n-i} & \sum_{i=1}^{n} t_{n-i}p_{n-i} \\ t_0 & t_1 & \cdots & t_{n-2} & t_{n-1} \\ t_{-1} & t_0 & \cdots & t_{n-3} & t_{n-2} \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ t_{-n+2} & t_{-n+3} & \cdots & t_0 & t_1 \end{bmatrix}$$

和

$$TC^{(2)}=\begin{bmatrix} t_1 & t_2 & \cdots & t_{n-1} & \sum_{j=1}^{n} t_{j-1}p_{j-1} \\ t_0 & t_1 & \cdots & t_{n-2} & \sum_{j=1}^{n} t_{j-2}p_{j-1} \\ t_{-1} & t_0 & \cdots & t_{n-3} & \sum_{j=1}^{n} t_{j-3}p_{j-1} \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ t_{-n+2} & t_{-n+3} & \cdots & t_0 & \sum_{j=1}^{n} t_{j-n}p_{j-1} \end{bmatrix}.$$

从而,欲使(3.11)成立,只须证明存在 $p_{n-1},p_{n-2},\cdots,p_0$,满足

$$\sum_{i=1}^{n} t_{k-i}p_{n-i}=t_k,\ \sum_{j=1}^{n} t_{j-k}p_{j-1}=t_{n+1-k},\ k=1,\cdots,n. \tag{3.15}$$

这里,注意到 $\sum_{i=1}^{n} t_{n-i}p_{n-i}=\sum_{j=1}^{n} t_{j-1}p_{j-1}$,令

$$t_n\equiv\sum_{i=1}^{n} t_{n-i}p_{n-i}=\sum_{j=1}^{n} t_{j-1}p_{j-1}, \tag{3.16}$$

(3.15)是关于 $p_{n-1},p_{n-2},\cdots,p_0$ 的两个线性方程组,可以表示成

$$\left(p_{n-1},p_{n-2},\cdots,p_0\right)T=\left(t_1,t_2,\cdots,t_n\right) \tag{3.17}$$

和

$$T\left(p_0,\cdots,p_{n-2},p_{n-1}\right)^{\mathrm{T}}=\left(t_n,\cdots,t_2,t_1\right)^{\mathrm{T}}, \tag{3.18}$$

由于(3.18)可以改写成

$$\left(p_0,\cdots,p_{n-2},p_{n-1}\right)T^{\mathrm{T}}=\left(t_n,\cdots,t_2,t_1\right),$$

进一步,有

$$\left[\left(p_0,\cdots,p_{n-1},p_{n-1}\right)E\right]\left(ET^{\mathrm{T}}E\right)=\left(t_n,\cdots,t_2,t_1\right)E,$$

由此及(3.14)得知,(3.18)等价于(3.17).因为 T 非奇异,所以(3.17)对于任意取定的 t_n 有唯一解.

反之,如果成立(3.11),那么从(3.12)和(3.13)可得

$$t_{ij}=t_{i-1,j-1},\quad i,j=2,\cdots,n.$$

由此并令

$$t_{j-1}\equiv t_{1j},\ t_{1-i}\equiv t_{i1},\quad i,j=1,\cdots,n,$$

推出$T=[t_{ij}]$形如(3.2). □

6.4 Hankel 矩阵

6.4.1 定义 $H=[h_{ij}]\in\mathbb{C}^{n\times n}$称为 **Hankel 矩阵**,如果满足

$$h_{ij}=h_{i+j-1},\quad i,j=1,\cdots,n. \tag{4.1}$$

具体地,形如

$$H=\begin{bmatrix} h_1 & h_2 & h_3 & \cdots & h_n \\ h_2 & h_3 & & h_n & h_{n+1} \\ h_3 & & \iddots & \iddots & h_{n+2} \\ \vdots & h_n & h_{n+1} & \iddots & \vdots \\ h_n & h_{n+1} & h_{n+2} & \cdots & h_{2n-1} \end{bmatrix}. \tag{4.2}$$

Hankel 矩阵是对称矩阵,完全由其第1行和第n列的$2n-1$个元素

$$h_1,h_2,\cdots,h_n,h_{n+1},\cdots,h_{2n-1}\in\mathbb{C}$$

确定.

Hankel 矩阵和 Toeplitz 矩阵有明显而紧密的联系,写成如下定理(证明留作练习).

6.4.2 定理 若$H\in\mathbb{C}^{n\times n}$是 Hankel 矩阵,则$EH$和$HE$是 Toeplitz

矩阵.若$T \in \mathbb{C}^{n\times n}$是 Toeplitz 矩阵,则$ET$和$TE$是 Hankel 矩阵.其中,$E = [e_n \quad \cdots \quad e_1] \in \mathbb{R}^{n\times n}$是反序单位矩阵.

根据这一性质,对于 Toeplitz 矩阵和 Hankel 矩阵,有关两者之一的结果,常可直接应用于其二.下面的 **6.4.3** 平行于 **6.3.4** 即为一例.

因而,有时将 Toeplitz 矩阵和 Hankel 矩阵统称为**条纹矩阵**(striped matrix).

条纹矩阵出现在许多应用领域中,包括随机过程、时序分析、数字滤波、幂矩量等.

6.4.3 定理 非奇异矩阵$H \in \mathbb{C}^{n\times n}$为 Hankel 矩阵的充分必要条件是存在$n\times n$友矩阵

$$C = \begin{bmatrix} 0 & 1 & 0 & \cdots & 0 \\ \vdots & \ddots & \ddots & \ddots & \vdots \\ \vdots & & \ddots & \ddots & 0 \\ 0 & \cdots & \cdots & 0 & 1 \\ p_0 & p_1 & \cdots & p_{n-2} & p_{n-1} \end{bmatrix}, \tag{4.3}$$

使得

$$CH = HC^{\mathrm{T}}. \tag{4.4}$$

证 依 **6.4.2**, H 为非奇异 Hankel 矩阵的充分必要条件是EH和HE为非奇异 Toeplitz 矩阵.$E = [e_n \quad \cdots \quad e_1] \in \mathbb{R}^{n\times n}$是反序单位矩阵.

再依 **6.3.4**, 非奇异矩阵$T \in \mathbb{C}^{n\times n}$为 Toeplitz 矩阵的充分必要条件是存在形如(3.9)和(3.10)的$n\times n$友矩阵$C^{(1)}$和$C^{(2)}$,使得

$$C^{(1)}T = TC^{(2)}.$$

现取$T = EH$,于是

$$C^{(1)}EH = EHC^{(2)},$$

从而

$$\left(EC^{(1)}E\right)H = HC^{(2)}.$$

对于此式,令

$$C \equiv EC^{(1)}E,$$

易见 C 形如(4.3),而 $C^{(2)} = C^{\mathrm{T}}$;因此即成为(4.4). □

6.4.4 定义 设 $H = [h_{i+j-1}] \in \mathbb{C}^{n\times n}$ 是 Hankel 矩阵,如果其组成元素成一算术序列:

$$h_k = a + (k-1)d,\quad k = 1,\cdots,2n-1, \tag{4.5}$$

则称 H 为**算术矩阵**(arithmetic matrix).具体地,形如

$$H = \begin{bmatrix} a & a+d & a+2d & \cdots & a+(n-1)d \\ a+d & a+2d & a+3d & \cdots & a+nd \\ a+2d & a+3d & \cdots & \cdots & a+(n+1)d \\ a+3d & \cdots & \cdots & \cdots & a+(n+2)d \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ a+(n-1)d & a+nd & a+(n+1)d & \cdots & a+(2n-2)d \end{bmatrix}. \tag{4.6}$$

6.4.5 引理 设 $H = [h_{i+j-1}] \in \mathbb{C}^{n\times n}$ 是形如(4.6)的算术矩阵,而且 $d \neq 0$,则 $\mathrm{rank}H = 2$.

证 因为当 $d \neq 0$ 时,H 的2阶主子式

$$\begin{vmatrix} a & a+d \\ a+d & a+2d \end{vmatrix} = \begin{vmatrix} a & d \\ a+d & d \end{vmatrix} = ad - (a+d)d = -d^2 \neq 0,$$

所以 $\mathrm{rank}H \geq 2$.

另一方面,容易验证 H 的任意三行线性相关,因此又有 $\mathrm{rank}H \leq 2$. □

6.4.6 定理 设 $H \in \mathbb{C}^{n\times n}$,$\mathrm{rank}H = 2$,且 $I + H$ 是非奇异的,则

$$(I+H)^{-1} = I + \frac{H^2 - (1+\sigma_1)H}{1+\sigma_1+\sigma_2}, \tag{4.7}$$

其中

$$\sigma_1 = \mathrm{tr}H, \quad \sigma_2 = \frac{1}{2}\left((\mathrm{tr}H)^2 - \mathrm{tr}H^2\right). \tag{4.8}$$

当 H 是算术矩阵(4.6)时,

$$\sigma_1 = na + n(n-1)d, \quad \sigma_2 = -n^2(n^2-1)d^2/12. \tag{4.9}$$

证 依 **5.1.2**,存在非奇异矩阵 $P \in \mathbb{C}^{n\times n}$,使得

$$P^{-1}HP = J, \tag{4.10}$$

其中 J 是 H 的 Jordan 标准形.

因为 $\mathrm{rank}H = 2$, J 必形如

$$J = \begin{bmatrix} \lambda_1 & & & \\ & \lambda_2 & & \\ & & 0 & \\ & & & \ddots & \\ & & & & 0 \end{bmatrix} \in \mathbb{C}^{n\times n} \tag{4.11}$$

或者

$$J = \begin{bmatrix} \lambda & & & \\ & \lambda & & \\ & & 0 & \\ & & & \ddots & \\ & & & & 0 \end{bmatrix} \in \mathbb{C}^{n\times n}, \tag{4.12}$$

这里, $\lambda_1, \lambda_2, \lambda \in \mathbb{C}$,且均不为零.

先考虑 J 形如(4.11).从(4.10),有

$$I + H = P(I+J)P^{-1}.$$

由此,并依假设 $I+H$ 非奇异,得

$$(I+H)^{-1} = P(I+J)^{-1}P$$

$$
= P \begin{bmatrix} \frac{1}{1+\lambda_1} & & & & 0 \\ & \frac{1}{1+\lambda_2} & & & \\ & & 1 & & \\ & & & \ddots & \\ 0 & & & & 1 \end{bmatrix} P^{-1}
$$

$$
= P \begin{bmatrix} 1 - \frac{\lambda_1(1+\lambda_2)}{(1+\lambda_1)(1+\lambda_2)} & & & & 0 \\ & 1 - \frac{\lambda_2(1+\lambda_1)}{(1+\lambda_1)(1+\lambda_2)} & & & \\ & & 1 & & \\ & & & \ddots & \\ 0 & & & & 1 \end{bmatrix} P^{-1}
$$

$$
= P\left(I + \frac{J^2 - (1+\sigma_1)J}{1+\sigma_1+\sigma_2} \right) P^{-1} = I + \frac{H^2 - (1+\sigma_1)H}{1+\sigma_1+\sigma_2},
$$

其中

$$
\sigma_1 \equiv \lambda_1 + \lambda_2 = \mathrm{tr} H , \ \sigma_2 \equiv \lambda_1 \lambda_2 = \frac{1}{2}\left((\mathrm{tr} H)^2 - \mathrm{tr} H^2 \right).
$$

再考虑 J 形如(4.12).类似地,有

$$
(I+H)^{-1} = P \begin{bmatrix} \frac{1}{1+\lambda} & -\frac{1}{(1+\lambda)^2} & & & 0 \\ & \frac{1}{1+\lambda} & & & \\ & & 1 & & \\ & & & \ddots & \\ 0 & & & & 1 \end{bmatrix} P^{-1}
$$

$$
= P\begin{bmatrix} 1-\frac{\lambda(1+\lambda)}{(1+\lambda)^2} & -\frac{1}{(1+\lambda)^2} & & & 0 \\ & 1-\frac{\lambda(1+\lambda)}{(1+\lambda)^2} & & & \\ & & 1 & & \\ & 0 & & \ddots & \\ & & & & 1 \end{bmatrix} P^{-1}.
$$

由此可得 $(I+H)^{-1}$ 同样具有(4.7)的形式,此时

$$
\sigma_1 \equiv 2\lambda = \mathrm{tr}H, \quad \sigma_2 \equiv \lambda^2 = \frac{1}{2}\left((\mathrm{tr}H)^2 - \mathrm{tr}H^2\right).
$$

当 H 是算术矩阵(4.6)时,

$$
\begin{aligned}
\sigma_1 &= \mathrm{tr}H = a + (a+2d) + (a+4d) + \cdots + (a+(2n-2)d) \\
&= na + n(n-1)d. \\
\sigma_2 &= \frac{1}{2}\left((\mathrm{tr}H)^2 - \mathrm{tr}H^2\right) \\
&= \frac{1}{2}\left[(na+n(n-1)d)^2 - \sum_{j=0}^{n-1}\sum_{i=j}^{n+j-1}(a+id)^2\right] \\
&= -\frac{1}{12}n^2(n^2-1)d^2.
\end{aligned}
$$

□

6.4.7 定义 设 $H = [h_{i+j-1}] \in \mathbb{C}^{n\times n}$ 是 Hankel 矩阵,如果其组成元素成一几何序列

$$
h_1 = 1; \; h_k = \lambda^{k-1}, \; k = 2, \cdots, 2n-1, \tag{4.13}
$$

则称 H 为**几何矩阵**(geometric matrix).

6.4.8 定理 设 $H = [h_{i+j-1}] \in \mathbb{C}^{n\times n}$ 是由(4.13)确定的几何矩阵,则

(1) $\operatorname{rank}H = 1$.

(2) 当 $\lambda \neq \pm 1$ 时,

$$(I+H)^{-1} = I - \frac{(1-\lambda^2)H}{2-\lambda^2-\lambda^{2n}}. \tag{4.14}$$

证 (1) 显然, $\operatorname{rank}H \geq 1$, 又 H 的任两行是线性相关的.

(2) 根据(1), H 的 Jordan 标准形形如

$$P^{-1}HP = J = \begin{bmatrix} \mu & & & \\ & 0 & & \\ & & \ddots & \\ & & & 0 \end{bmatrix} \in \mathbb{C}^{n\times n},$$

其中 $P \in \mathbb{C}^{n\times n}$ 是非奇异矩阵.注意到

$$\mu = \operatorname{tr}H = 1+\lambda^2+\lambda^4+\cdots+\lambda^{2n-2} = \frac{1-\lambda^{2n}}{1-\lambda^2},$$

从而

$$(I+J)^{-1} = \begin{bmatrix} \frac{1}{1+\mu} & & & \\ & 1 & & \\ & & \ddots & \\ & & & 1 \end{bmatrix}$$

$$= \begin{bmatrix} 1-\frac{(1-\lambda^2)\mu}{2-\lambda^2-\lambda^{2n}} & & & \\ & 1 & & \\ & & \ddots & \\ & & & 1 \end{bmatrix}.$$

因此,从

$$(I+H)^{-1} = P(I+J)^{-1}P^{-1},$$

即得(4.14). □

6.4.9 定义　设 $H=[h_{i+j-1}]\in\mathbb{C}^{n\times n}$ 是 Hankel 矩阵,如果其组成元素成一调和序列

$$h_k=\frac{1}{a+(k-1)d},\ k=1,\cdots,2n-1, \tag{4.15}$$

则称 H 为**调和矩阵**(harmonic matrix).这里假定 a 和 d 使得每个 h_k 的分母不为零.

特别,当 $a=d=1$ 时,调和矩阵就是 Hilbert 矩阵(见 **2.5.5**).

6.4.10 定理　设 $H=[h_{i+j-1}]\in\mathbb{C}^{n\times n}$ 是由(4.15)确定的调和矩阵,则

$$\det H=\frac{\prod_{k=1}^{n}\left[(n-k)!\,d^{n-k}\right]^2}{\prod_{k=1}^{n}D_k}, \tag{4.16}$$

其中 D_k 是 H 的第 k 行中的各分母之积,

$$D_k=\prod_{j=k-1}^{n+k-2}(a+jd),\ k=1,\cdots,n. \tag{4.17}$$

证明从略.

6.4.11 定义　$L=[l_{ij}]\in\mathbb{C}^{n\times n}$ 称为 **Loewner 矩阵**,如果

$$l_{ij}=\frac{c_i-d_j}{y_i-z_j},\ i,j=1,\cdots,n, \tag{4.18}$$

其中 $c_1,\cdots,c_n$ 和 $d_1,\cdots,d_n$ 可以是任何数;$y_1,\cdots,y_n$ 和 $z_1,\cdots,z_n$ 则是保证所有 l_{ij} 的分母不为零的任何数.

如此矩阵在有理插值中有用.

6.4.12 定理 设 $W(t) \equiv W(t_1, \cdots, t_n) \in \mathbb{C}^{n \times n}$ 的第 i 行是由 $n-1$ 次多项式

$$f_i(\lambda) \equiv \prod_{j=1, j \neq i}^{n} (\lambda - t_j) \tag{4.19}$$

按 λ 之幂递升次序的系数组成的, $i = 1, \cdots, n$, 则 Hankel 矩阵和 Loewner 矩阵有如下关系:

(1) 如果 $H \in \mathbb{C}^{n \times n}$ 是任一 Hankel 矩阵,那么

$$L \equiv W(y) H (W(z))^{\mathrm{T}} \tag{4.20}$$

是 Loewner 矩阵,其中 $y = (y_1, \cdots, y_n)^{\mathrm{T}}, z = (z_1, \cdots, z_n)^{\mathrm{T}}$.

(2) 反之,在(4.20)中,如果 L 是 Loewner 矩阵,那么 H 是 Hankel 矩阵,并且

$$H = V(y) D^{-1}(y) L D^{-1}(z) (V(z))^{\mathrm{T}}, \tag{4.21}$$

其中

$$D(t) = \operatorname{diag}(f_1(t_1), \cdots, f_n(t_n)), \tag{4.22}$$

而 $V(y), V(z) \in \mathbb{C}^{n \times n}$ 是 Vandermode 矩阵(见 **2.6.1**).

证明从略.

6.5 若干其它条纹矩阵

6.5.1 定义 $A = [a_{ij}] \in \mathbb{R}^{n \times n}$ 称为 **Pei 矩阵**,如果

$$a_{ij} = \begin{cases} d, & i = j, \\ 1, & i \neq j, \end{cases} \quad i, j = 1, \cdots, n. \tag{5.1}$$

如此矩阵是 Toeplitz 矩阵的特殊情形.

6.5.2 定理 设 $A \in \mathbb{R}^{n\times n}$ 是由(5.1)确定的 Pei 矩阵,则

(1) $\det A = (d-1)^{n-1}(d+n-1)$.

(2) 当 $d \neq 1$ 和 $d \neq -n+1$ 时, $A^{-1} = [\alpha_{ij}]$,

$$\alpha_{ij} = \begin{cases} \dfrac{d+n-2}{d(d+n-2)-(n-1)}, & i=j, \\ \dfrac{-1}{d(d+n-2)-(n-1)}, & i \neq j, \end{cases} \quad i,j = 1,\cdots,n.$$

(3) A 的特征值为

$$\lambda_1 = d+n-1, \quad \lambda_2 = \cdots = \lambda_n = d-1.$$

相应的正交规范特征向量组为

$$x^{(1)} = \frac{1}{\sqrt{n}}(1,\cdots,1)^{\mathrm{T}},$$

$$x^{(k)} = \frac{1}{\sqrt{k(k-1)}}(\underbrace{1,\cdots,1}_{k-1},-k+1,\underbrace{0,\cdots,0}_{n-k})^{\mathrm{T}}, \quad k = 2,\cdots,n.$$

(4) A 是对称的且当 $d > 1$ 时是正定的;此时, A 关于矩阵求逆的条件数

$$\kappa(A) = \frac{\lambda_1}{\lambda_2} = \frac{d+n-1}{d-1}.$$

证 A 的特征多项式

$$\det(A-\lambda I) = \begin{vmatrix} d-\lambda & 1 & 1 & \cdots & 1 \\ 1 & d-\lambda & 1 & \cdots & 1 \\ 1 & 1 & d-\lambda & \cdots & 1 \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ 1 & 1 & 1 & \cdots & d-\lambda \end{vmatrix}$$

$$
=\begin{vmatrix}
d+n-1-\lambda & 1 & 1 & \cdots & 1 \\
d+n-1-\lambda & d-\lambda & 1 & \cdots & 1 \\
d+n-1-\lambda & 1 & d-\lambda & \cdots & 1 \\
\cdots & \cdots & \cdots & \cdots & \cdots \\
d+n-1-\lambda & 1 & 1 & \cdots & d-\lambda
\end{vmatrix}
$$

$$
=\begin{vmatrix}
d=n-1-\lambda & 1 & 1 & \cdots & 1 \\
0 & d-1-\lambda & 0 & \cdots & 0 \\
0 & 0 & d-1-\lambda & \cdots & 0 \\
\cdots & \cdots & \cdots & \cdots & \cdots \\
0 & 0 & 0 & \cdots & d-1-\lambda
\end{vmatrix}
$$

$$
=(d+n-1-\lambda)(d-1-\lambda)^{n-1}, \tag{5.2}
$$

推出结论(3)中所示的特征值;从而,容易计算出(3)中所示的相应的正交规范特征向量组.显然,(5.2)中取 $\lambda=0$ 时即得结论(1).

现在,令 P 是以(3)中的特征向量为列而构成的正交矩阵,

$$
P=\begin{bmatrix} x^{(1)} & \cdots & x^{(n)} \end{bmatrix}.
$$

令 D 是以(3)中的特征值为对角元素而构成的对角矩阵,

$$
D=\mathrm{diag}(\lambda_1,\cdots,\lambda_n).
$$

于是,因为有 $A=PDP^{-1}$,而且当 $d\neq 1$ 和 $d\neq -n+1$ 时,

$$
D^{-1}=\mathrm{diag}\left(\frac{1}{\lambda_1},\cdots,\frac{1}{\lambda_n}\right)
$$

$$
=\mathrm{diag}\left(\frac{1}{\mathrm{d+n-1}},\frac{1}{d-1},\cdots,\frac{1}{d-1}\right),
$$

所以当 $d\neq 1$ 和 $d\neq -n+1$ 时,

$$
A^{-1}=PD^{-1}P^{-1}.
$$

实现此等式右端的矩阵相乘,即得结论(2).

最后,由于 A 对称,而且根据(3),当 $d>1$ 时特征值全为正数,从而 A 正定.因此,再根据 **2.5.3** 的(3),即得结论(4). □

当 $d>1$ 而接近1时,从条件数

$$\kappa(A)=(d+n-1)/(d-1)$$

看出, Pei 矩阵严重病态.

6.5.3 定义 $A=[a_{ij}]\in\mathbb{R}^{n\times n}$ 称为 **Rodman 矩阵**,如果

$$a_{ij}=\begin{cases}1, & i=j,\\ t, & i\neq j,\end{cases}\quad i,j=1,\cdots,n. \tag{5.3}$$

对于 Rodman 矩阵 A:当 $t=0$ 时,即为单位矩阵,可以不考虑;而当 $t\neq 0$ 时,令 $d=1/t$,则

$$\hat{A}\equiv dA$$

是 Pei 矩阵.因此,通过对 $\hat{A}$ 应用 **6.5.2**,便可得到关于 Rodman 矩阵 $A=t\hat{A}$ 的下述对应结果.

6.5.4 定理 设 $A\in\mathbb{R}^{n\times n}$ 是由(5.3)确定的 Rodman 矩阵,则

(1) $\det A=t^n\left(\frac{1}{t}-1\right)^{n-1}\left(\frac{1}{t}+n-1\right)$.

(2) 当 $t\neq 1$ 和 $t\neq\frac{1}{-n+1}$ 时, $A^{-1}=[\alpha_{ij}]$,

$$\alpha_{ij}=\begin{cases}\dfrac{1+t(n-2)}{1+t(n-2)-t^2(n-1)}, & i=j,\\[2ex] \dfrac{-t}{1+t(n-2)-t^2(n-1)}, & i\neq j,\end{cases}\quad i,j=1,\cdots,n.$$

(3) A 的特征值为

$$\lambda_1=1+nt-t,\ \lambda_2=\cdots=\lambda_n=1-t.$$

相应的特征向量分别是 $x^{(1)}=(1,\cdots,1)^T$ 以及满足方程

$$x_1 + \cdots + x_n = 0$$

的任何 $n-1$ 个线性无关的向量.

(4) A 是对称的且当 $t<1$ 时是正定的.

证明留作练习.

6.5.5 定义 $A=[a_{ij}]\in\mathbb{R}^{n\times n}$ 称为 **Lotkin 矩阵**,如果

$$a_{ij}=\begin{cases} 1, & i=1, \quad j=1,\cdots,n, \\ \dfrac{1}{i+j-1}, & i=2,\cdots,n, \quad j=1,\cdots,n. \end{cases} \tag{5.4}$$

Lotkin 矩阵相当于将 Hilbert 矩阵(见 **2.5.5**)的第 1 行的所有元素修改为 1 而成的.

6.5.6 定理 设 $A\in\mathbb{R}^{n\times n}$ 是由(5.4)确定的 Lotkin 矩阵,则

(1) $\det A=(-1)^{n-1}\delta_n^{-1}$, δ_n 由如下递推公式产生:

$$\delta_1=1, \quad \delta_{k+1}=\binom{2k}{k-1}\binom{2k}{k}(2k+1)\delta_k, \quad k=1,\cdots,n-1.$$

(2) $A^{-1}=[\alpha_{ij}]$,

$$\alpha_{i1}=(-1)^{n-i}\binom{n+i-1}{i-1}\binom{n}{i},$$

$$\alpha_{i,j+1}=(-1)^{i-j}i\binom{i+j}{j}\binom{i+j-1}{j-1}\binom{n+i-1}{i+j}\binom{n+j}{i+j}$$

$$=\frac{(-1)^{i-j}(n+i-1)!(n+j)!}{ij(i+j)[(i-1)!(j-1)!]^2(n-i)!(n-j-1)!},$$

$$i=1,\cdots,n, \quad j=1,\cdots,n-1.$$

证明从略.

当 $n=10$ 时, $\det A=\left(46207\times 10^{47}\right)^{-1}$, A 的最大特征值和最小特征值为

$$\lambda_{\max}=2.428554$$

和

$$\lambda_{\min}=-0.1267649\times 10^{-12}.$$

$B\equiv A^{\mathrm{T}}A$ 是对称正定的,但其条件数等性态比 A 更坏.

6.5.7 定义 W.L.Frank 给出矩阵

$$F_1=\begin{bmatrix} 1 & 1 & & & & & \\ 1 & 2 & 2 & & & \text{\Large 0} & \\ 1 & 2 & 3 & 3 & & & \\ 1 & 2 & 3 & 4 & \ddots & & \\ \vdots & \vdots & \vdots & \vdots & \ddots & \ddots & \\ 1 & 2 & 3 & 4 & \cdots & n-1 & n-1 \\ 1 & 2 & 3 & 4 & \cdots & n-1 & n \end{bmatrix}\in\mathbb{R}^{n\times n} \quad (5.5)$$

和

$$F_2=\begin{bmatrix} 12 & 11 & 10 & 9 & \cdots & 2 & 1 \\ 11 & 11 & 10 & 9 & \cdots & 2 & 1 \\ 10 & 10 & 10 & 9 & \cdots & 2 & 1 \\ 9 & 9 & 9 & 9 & \cdots & 2 & 1 \\ \vdots & \vdots & \vdots & \vdots & & \vdots & \vdots \\ 2 & 2 & 2 & 2 & \cdots & 2 & 1 \\ 1 & 1 & 1 & 1 & \cdots & 1 & 1 \end{bmatrix}\in\mathbb{R}^{12\times 12}, \quad (5.6)$$

F_1 和 F_2 用作特征值计算的测试矩阵.

容易证明(留作练习):

$$\det F_1 = 1, \quad \forall n = 1,2,\cdots \tag{5.7}$$

F_1 的行列式对其左下角附近元素的改变很敏感.

F_2 的特征值为

$$\lambda_i = \frac{1}{2}\left(1-\cos\frac{(2i-1)\pi}{25}\right)^{-1}, \quad i = 1,\cdots,12. \tag{5.8}$$

容易看出, EF_2E 属于下面定义的 Brownian 矩阵, 这里 $E = [e_n \quad \cdots \quad e_1]$ 是反序单位矩阵.

6.5.8 定义 $B = [b_{ij}] \in \mathbf{C}^{n\times n}$ 称为 **Brownian 矩阵**, 如果

$$\begin{array}{ll} b_{i,j+1} = b_{ij}, & j > i, \\ b_{i+1,j} = b_{ij}, & j < i, \end{array} \quad i, j = 1,\cdots,n-1 \tag{5.9}$$

具体地, 形如

$$B = \begin{bmatrix} b_1 & b_{n+1} & b_{n+1} & \cdots & b_{n+1} & b_{n+1} \\ b_{2n} & b_2 & b_{n+2} & \cdots & b_{n+2} & b_{n+2} \\ b_{2n} & b_{2n+1} & b_3 & \cdots & b_{n+3} & b_{n+3} \\ \vdots & \vdots & \vdots & & \cdots & \cdots \\ b_{2n} & b_{2n+1} & b_{2n+2} & \vdots & b_{n-1} & b_{2n-1} \\ b_{2n} & b_{2n+1} & b_{2n+2} & \vdots & b_{3n-2} & b_n \end{bmatrix}. \tag{5.10}$$

Brownian 矩阵完全由包含主对角线的三对角带上的 $3n-2$ 个元素确定.

6.5.9 定理 设 $B \in \mathbf{C}^{n\times n}$ 是任一 Brownian 矩阵, 则 B 相合于一个三对角矩阵 $A \in \mathbf{C}^{n\times n}$,

$$A = PBP^{\mathrm{T}}$$

其中 $P=[p_{ij}]$ 是双对角 Toeplitz 矩阵,

$$p_{ij}=\begin{cases}1, & j=i, \quad i=1,\cdots,n,\\ -1, & j=i-1, \quad i=2,\cdots,n,\\ 0, & j\neq i,i-1.\end{cases}$$

证明留作练习.

6.5.10 定义 设 $G=[g_{ij}]\in\mathbb{R}^{n\times n}$,

$$g_{ij}=2\min\{i,j\}-1,\ i,j=1,\cdots,n,$$

G 称为 **Givens 矩阵**.具体地,

$$G=\begin{bmatrix}1 & 1 & 1 & \cdots & 1\\ 1 & 3 & 3 & \cdots & 3\\ 1 & 3 & 5 & \cdots & 5\\ \vdots & \vdots & \vdots & & \cdots\\ 1 & 3 & 5 & \vdots & 2n-1\end{bmatrix}.$$

显然, G 是 Brownian 矩阵.

可以算出,

$$G^{-1}=\frac{1}{2}\begin{bmatrix}3 & -1 & & & & \\ -1 & 2 & -1 & & \text{0} & \\ & -1 & 2 & -1 & & \\ & & \ddots & \ddots & \ddots & \\ & \text{0} & & -1 & 2 & -1\\ & & & & -1 & 1\end{bmatrix},$$

G 和 G^{-1} 都是对称正定的. G 的条件数

$$\kappa(A)\approx 16n^2/\pi^2,$$

因此,稍大的n,G便是病态的.

6.6 中心对称矩阵和中心斜对称矩阵

6.6.1 定义 $C=[c_{ij}]\in\mathbf{C}^{n\times n}$称为**中心对称矩阵**(centrosymmetric matrix),如果

$$c_{ij}=c_{n+1-i,n+1-j},\quad i,j=1,\cdots,n. \tag{6.1}$$

具体地,形如

$$C=\begin{bmatrix} c_{11} & \cdots & c_{1m} & c_{1,m+1} & \cdots & c_{1n} \\ \vdots & \ddots & \vdots & \vdots & ⋰ & \vdots \\ c_{m1} & \cdots & c_{mm} & c_{m,m+1} & \cdots & c_{mn} \\ c_{mn} & \cdots & c_{m,m+1} & c_{mm} & \cdots & c_{m1} \\ \vdots & ⋰ & \vdots & \vdots & \ddots & \vdots \\ c_{1n} & \cdots & c_{1,m+1} & c_{1m} & \cdots & c_{11} \end{bmatrix},n=2m, \tag{6.2}$$

或者

$$C=\begin{bmatrix} c_{11} & \cdots & c_{1m} & c_{1,m+1} & c_{1,m+2} & \cdots & c_{1n} \\ \vdots & \ddots & \vdots & \vdots & \vdots & ⋰ & \vdots \\ c_{m1} & \cdots & c_{mm} & c_{m,m+1} & c_{m,m+2} & \cdots & c_{mn} \\ c_{m+1,1} & \cdots & c_{m+1,m} & c_{m+1,m+1} & c_{m+1,m} & \cdots & c_{m+1,1} \\ c_{mn} & \cdots & c_{m,m+2} & c_{m,m+1} & c_{mm} & \cdots & c_{m1} \\ \vdots & ⋰ & \vdots & \vdots & \vdots & \ddots & \vdots \\ c_{1n} & \cdots & c_{1,m+2} & c_{1,m+1} & c_{1m} & \cdots & c_{11} \end{bmatrix},$$

$$n=2m+1. \tag{6.3}$$

中心对称矩阵,当 $n=2m$ 是偶数时,有 $n^2/2$ 个独立元素,当 $n=2m+1$ 是奇数时,有 $(n^2+1)/2$ 个独立元素.倒数第 r 行的元素是正数第 r 行的元素按反序排列而成.

6.6.2 定理 设 $C\in\mathbb{C}^{n\times n}$ 是中心对称矩阵,则

(1) 等价地

$$C=ECE, \tag{6.4}$$

其中 $E=\begin{bmatrix}e_n & \cdots & e_1\end{bmatrix}$ 是反序单位矩阵.

(2) 当 $n=2m$ 时, C 可以表示成

$$C=\begin{bmatrix}C_1 & C_2E\\ EC_2 & EC_1E\end{bmatrix}=\begin{bmatrix}I & 0\\ 0 & E\end{bmatrix}\begin{bmatrix}C_1 & C_2\\ C_2 & C_1\end{bmatrix}\begin{bmatrix}I & 0\\ 0 & E\end{bmatrix}, \tag{6.5}$$

其中 $C_1,C_2,I\in\mathbb{C}^{m\times m}$, $E=\begin{bmatrix}e_m & \cdots & e_1\end{bmatrix}\in\mathbb{C}^{m\times m}$.

(3) 当 $n=2m+1$ 是奇数时, C 可以表示成

$$C=\begin{bmatrix}C_1 & a & C_2E\\ b^T & \alpha & b^TE\\ EC_2 & Ea & EC_1E\end{bmatrix}$$

$$=\begin{bmatrix}I & 0 & 0\\ 0 & 1 & 0\\ 0 & 0 & E\end{bmatrix}\begin{bmatrix}C_1 & a & C_2\\ b^T & \alpha & b^T\\ C_2 & a & C_1\end{bmatrix}\begin{bmatrix}I & 0 & 0\\ 0 & 1 & 0\\ 0 & 0 & E\end{bmatrix}, \tag{6.6}$$

其中 $E=\begin{bmatrix}e_m & \cdots & e_1\end{bmatrix}$, $C_1,C_2,I\in\mathbb{C}^{m\times m}$; $a,b\in\mathbb{C}^m$; $\alpha\in\mathbb{C}$.

证明留作练习.

6.6.3 定理 设 $C\in\mathbb{C}^{2m\times 2m}$ 是形如(6.5)的中心对称矩阵,并且令

$$X=C_1+C_2,\quad Y=C_1-C_2,$$

和

$$C_3=\frac{1}{2}\left(X^{-1}+Y^{-1}\right),\quad C_4=\frac{1}{2}\left(X^{-1}-Y^{-1}\right),$$

则

(1) $\det C = \det X \det Y$.

(2) $\lambda(C) = \lambda(X) \cup \lambda(Y)$,即 C 的特征值是由 X 的特征值和 Y 的特征值组成的.

(3) $C^{-1} = \begin{bmatrix} C_3 & C_4 E \\ EC_4 & EC_3 E \end{bmatrix}$.

证 取 $I = [e_1 \quad \cdots \quad e_m], E = [e_m \quad \cdots \quad e_1] \in \mathbf{C}^{m\times m}$,

$$T = \begin{bmatrix} I & -I \\ E & E \end{bmatrix},$$

则 $T^{-1} = \frac{1}{2} T^{T}$,于是

$$T^{-1}CT = \frac{1}{2}\begin{bmatrix} I & E \\ -I & E \end{bmatrix}\begin{bmatrix} I & 0 \\ 0 & E \end{bmatrix}\begin{bmatrix} C_1 & C_2 \\ C_2 & C_1 \end{bmatrix}\begin{bmatrix} I & 0 \\ 0 & E \end{bmatrix}\begin{bmatrix} I & -I \\ E & E \end{bmatrix}$$

$$= \begin{bmatrix} C_1 + C_2 & 0 \\ 0 & C_1 - C_2 \end{bmatrix} = \begin{bmatrix} X & 0 \\ 0 & Y \end{bmatrix},$$

由此,直接得出(1)和(2).而且

$$C^{-1} = T\begin{bmatrix} X^{-1} & 0 \\ 0 & Y^{-1} \end{bmatrix}T^{-1} = \frac{1}{2}\begin{bmatrix} I & -I \\ E & E \end{bmatrix}\begin{bmatrix} X^{-1} & 0 \\ 0 & Y^{-1} \end{bmatrix}\begin{bmatrix} I & E \\ -I & E \end{bmatrix}$$

$$= \frac{1}{2}\begin{bmatrix} X^{-1} + Y^{-1} & (X^{-1} - Y^{-1})E \\ E(X^{-1} - Y^{-1}) & E(X^{-1} + Y^{-1})E \end{bmatrix}$$

$$= \begin{bmatrix} C_3 & C_4 E \\ EC_4 & EC_3 E \end{bmatrix}.$$ □

6.6.4 定义 $S \in \mathbf{C}^{n\times n}$ 称为**中心斜对称矩阵**(centroskew symmetric matrix),如果

$$S = -ESE, \tag{6.7}$$

其中 $E=\begin{bmatrix}e_n & \cdots & e_1\end{bmatrix}$ 是反序单位矩阵.

6.6.5 定理 设 $S\in\mathbf{C}^{(2m+1)\times(2m+1)}$ 是中心斜对称矩阵,则

(1) S 可表示成

$$S=\begin{bmatrix} S_1 & a & -S_2E \\ b^{\mathrm{T}} & 0 & -b^{\mathrm{T}}E \\ ES_2 & -Ea & -ES_1E \end{bmatrix}, \tag{6.8}$$

其中 $E=\begin{bmatrix}e_m & \cdots & e_1\end{bmatrix}, S_1,S_2\in\mathbf{C}^{m\times m}; a,b\in\mathbf{C}^m$.

(2) $\det S=0$.

证 (1) 设

$$S=\begin{bmatrix} S_1 & a & S_3 \\ b^{\mathrm{T}} & \alpha & c^{\mathrm{T}} \\ ES_2 & d & S_4 \end{bmatrix},$$

其中 $E=\begin{bmatrix}e_m & \cdots & e_1\end{bmatrix}, S_1,S_2,S_3,S_4\in\mathbf{C}^{m\times m}; a,b,c,d\in\mathbb{C}^m$; $\alpha\in\mathbf{C}$. S 满足(6.7),得

$$S\equiv\begin{bmatrix} S_1 & a & S_3 \\ b^{\mathrm{T}} & \alpha & c^{\mathrm{T}} \\ ES_2 & d & S_4 \end{bmatrix}=-ESE$$

$$\equiv-\begin{bmatrix} 0 & 0 & E \\ 0 & 1 & 0 \\ E & 0 & 0 \end{bmatrix}\begin{bmatrix} S_1 & a & S_3 \\ b^{\mathrm{T}} & \alpha & c^{\mathrm{T}} \\ ES_2 & d & S_4 \end{bmatrix}\begin{bmatrix} 0 & 0 & E \\ 0 & 1 & 0 \\ E & 0 & 0 \end{bmatrix}$$

$$=-\begin{bmatrix} ES_4E & Ed & S_2E \\ c^{\mathrm{T}}E & \alpha & b^{\mathrm{T}}E \\ ES_3E & Ea & ES_1E \end{bmatrix},$$

由此推出

$$S_4=-ES_1E\,, S_3=-S_2E\,, c^{\mathrm{T}}=-b^{\mathrm{T}}E\,, d=-Ea\,, \alpha=0\,.$$

(2) 取 $I=\begin{bmatrix}e_1 & \cdots & e_m\end{bmatrix}, E=\begin{bmatrix}e_m & \cdots & e_1\end{bmatrix}\in\mathbf{C}^{m\times m}$,

$$V=\begin{bmatrix} I & 0 & I \\ 0 & 1 & 0 \\ E & 0 & -E \end{bmatrix},$$

则 $V^{-1}=\frac{1}{2}V^{\mathrm{T}}$,于是

$$\begin{aligned} & V^{-1}SV \\ & =\frac{1}{2}\begin{bmatrix} I & 0 & E \\ 0 & 1 & 0 \\ I & 0 & -E \end{bmatrix}\begin{bmatrix} S_1 & a & -S_2E \\ b^{\mathrm{T}} & 0 & -b^{\mathrm{T}}E \\ ES_2 & -Ea & -ES_1E \end{bmatrix}\begin{bmatrix} I & 0 & I \\ 0 & 1 & 0 \\ E & 0 & -E \end{bmatrix} \\ & =\begin{bmatrix} 0 & 0 & S_1+S_2 \\ 0 & 0 & b^{\mathrm{T}} \\ S_1-S_2 & a & 0 \end{bmatrix}, \end{aligned}$$

由此推出 $\det S=\det\left(V^{-1}SV\right)=0$. □

6.6.6 定义 $C\in\mathbf{C}^{n\times n}$ 称为**中心自伴矩阵**或**中心 Hermite 矩阵**,如果

$$C=EC^{*}E. \tag{6.9}$$

C 称为**中心斜自伴矩阵**或**中心斜 Hermite 矩阵**,如果

$$C=-EC^{*}E, \tag{6.10}$$

其中 $E=\begin{bmatrix}e_n & \cdots & e_1\end{bmatrix}$ 是反序单位矩阵.

容易证明(留作练习):

(1) 若 C 是中心自伴矩阵,则 C^{*} 也是中心自伴矩阵.

(2) 若 C 是中心斜自伴矩阵,则 $\pm iC$ 是中心自伴矩阵.

关于中心对称矩阵的进一步结果可参见本书参考文献中列出的有关资料.

6.7 同伴矩阵

6.7.1 定义 $A\in\mathbb{R}^{n\times n}$ 称为**同伴矩阵**(comrade matrix),如果 A 可以表示成

$$
A=\begin{bmatrix}
-\frac{\beta_1}{\alpha_1} & \frac{1}{\alpha_1} & & & & & \\
\frac{\gamma_2}{\alpha_2} & -\frac{\beta_2}{\alpha_2} & \frac{1}{\alpha_2} & & \text{\Large 0} & & \\
 & \frac{\gamma_3}{\alpha_3} & -\frac{\beta_3}{\alpha_3} & \frac{1}{\alpha_3} & & & \\
 & & \cdots & \cdots & \cdots & & \\
 & \text{\Large 0} & & \frac{\gamma_{n-1}}{\alpha_{n-1}} & -\frac{\beta_{n-1}}{\alpha_{n-1}} & \frac{1}{\alpha_{n-1}} \\
-\frac{a_n}{\alpha_n} & \cdots & \cdots & -\frac{a_3}{\alpha_n} & \frac{\gamma_n-a_2}{\alpha_n} & \frac{-\beta_n-a_1}{\alpha_n}
\end{bmatrix}
\tag{7.1}
$$

其中 $\alpha_i>0$, $i=1,\cdots,n$.

显然,当对所有 i,令

$$\alpha_i=1,\ \beta_i=\gamma_i=0$$

时,同伴矩阵就是友矩阵.

取名"同伴(comrade)矩阵"是摹仿"友(companion)矩阵"一词,从下一定理更可见其缘由.两种场合,矩阵都是"伴同"于一种特殊的多项式.

6.7.2 定理 设同伴矩阵 $A\in\mathbb{R}^{n\times n}$ 形如(7.1),则 A 的特征多项式

$$\det(\lambda I-A)=\frac{1}{\alpha_1\cdots\alpha_n}a(\lambda), \tag{7.2}$$

其中

$$a(\lambda)=p_n(\lambda)+a_1p_{n-1}(\lambda)+\cdots+a_{n-1}p_1(\lambda)+a_np_0(\lambda), \tag{7.3}$$

多项式 $p_0(\lambda),\cdots,p_n(\lambda)$ 按如下递推关系产生:

$$\begin{cases} p_0(\lambda)=1, p_1(\lambda)=\alpha_1\lambda+\beta_1, \\ p_i(\lambda)=(\alpha_i\lambda+\beta_i)p_{i-1}(\lambda)-\gamma_ip_{i-2}(\lambda), \quad i=2,\cdots,n. \end{cases} \tag{7.4}$$

证 令

$$D=\mathrm{diag}(\alpha_1,\cdots,\alpha_n),$$

则

$$\det(\lambda D-DA)=\alpha_1\cdots\alpha_n\det(\lambda I-A). \tag{7.5}$$

另一方面,将

$$\det(\lambda D-DA)$$

$$=\begin{bmatrix} \lambda\alpha_1+\beta_1 & -1 & & & \\ -\gamma_2 & \cdots & \cdots & & \mathbf{O} \\ & \cdots & \cdots & \cdots & \\ \mathbf{0} & & -\gamma_{n-1} & \lambda\alpha_{n-1}+\beta_{n-1} & -1 \\ a_n & \cdots & a_3 & -\gamma_n+a_2 & \lambda\alpha_n+a_1+\beta_n \end{bmatrix}$$

按其最后一行展开,可得

$$\det(\lambda D-DA)=a_n+a_{n-1}p_1(\lambda)+a_{n-2}p_2(\lambda)+\cdots$$
$$+(a_2-\gamma_n)p_{n-2}(\lambda)+(a_1+a_n\lambda+\beta_n)p_{n-1}(\lambda),$$

其中 $p_1(\lambda),\cdots,p_{n-1}(\lambda)$ 由(7.4)确定.由此并连同(7.5)推出(7.2)及(7.3). □

一般,若对所有 i, $\alpha_i>0$, $\gamma_i>0$,则由(7.4)确定的多项式

$p_0(\lambda),\cdots,p_n(\lambda)$构成一个正交集.具体地说,存在区间$[c,d]$和权函数$w$,

$$w(\lambda)\geq 0,\quad \forall\lambda\in[c,d],$$

使得

$$\int_c^d w(\lambda)\mathrm{d}\lambda>0 \tag{7.6}$$

和

$$\int_c^d w(\lambda)p_i(\lambda)p_j(\lambda)\mathrm{d}\lambda=\begin{cases}0, & j\neq i,\\ \tau_i>0, & j=i,\end{cases}$$
$$i,j=0,1,\cdots,n. \tag{7.7}$$

6.7.3 定义 如果对所有i,$\alpha_i=\gamma_i=1$,$\beta_i=0$,则称(7.1)确定的A为**同事矩阵**(colleague matrix).

同事矩阵A对应的(7.4)产生的多项式是由

$$\begin{cases}S_0(\lambda)=1,S_1(\lambda)=\lambda,\\ S_i(\lambda)=\lambda S_{i-1}(\lambda)-S_{i-2}(\lambda), & i=2,3,\cdots\end{cases} \tag{7.8}$$

定义的 **Chebyshev 多项式集**.

6.7.4 定义 矩阵

$$L=\begin{bmatrix}0 & f_1 & & & \\ & \ddots & \ddots & & 0\\ & & \ddots & \ddots & \\ 0 & & & 0 & f_{n-1}\\ l_1 & l_2 & \cdots & l_{n-1} & l_n\end{bmatrix}\in\mathbb{R}^{n\times n},$$

$$f_i\neq 0,\ i=1,\cdots,n-1 \tag{7.9}$$

称为 **Leslie 矩阵**.

Leslie 矩阵属于(7.1)中对所有i,$\gamma_i=\beta_i=0$的特殊情形.出现

在人口模型中.

6.7.5 定义 矩阵

$$A=\begin{bmatrix} 0 & 1 & & & \\ -b_n & \ddots & \ddots & \text{\Large 0} & \\ & -b_{n-1} & \ddots & \ddots & \\ \text{\Large 0} & & \ddots & 0 & 1 \\ & & & -b_2 & -b_1 \end{bmatrix}\in\mathbb{R}^{n\times n},$$

$$b_i\neq 0,\ i=1,\cdots,n \tag{7.10}$$

称为 **Schwarz 矩阵**.

显然,Schwarz 矩阵也是(7.1)的特殊情形,出现在稳定性理论和网络理论中.

6.7.6 定理 设 $A\in\mathbb{R}^{n\times n}$ 是 Schwarz 矩阵,形如(7.10),并设

$$H\equiv 2b_1^2\operatorname{diag}(0,\cdots,0,1),$$

则

(1) 矩阵对 (A,H) 满足可控制条件(见 **3.14.15**):

$$\operatorname{rank}\left[H\quad A^{\mathrm{T}}H\quad \left(A^{\mathrm{T}}\right)^2H\quad \cdots\quad \left(A^{\mathrm{T}}\right)^{n-1}H\right]=n.$$

(2) Lyapunov 方程(见 **3.14.6**)

$$GA+A^{\mathrm{T}}G=-H$$

的解是

$$G=\operatorname{diag}\left(b_1\cdots b_n,b_1\cdots b_{n-1},\cdots,b_1b_2,b_1\right).$$

(3) A 的惯性(见 **3.14.2**)是

$$i(A)=(k,n-k,0),$$

其中 k 是序列

$$b_1,b_1b_2,\cdots,b_1\cdots b_n$$

中负项的个数.

证明从略.可作练习.

6.7.7 定义 $n\times n$ 矩阵

$$R=\begin{bmatrix} -b_1 & \sqrt{b_2} & & & 0 \\ -\sqrt{b_2} & 0 & \ddots & & \\ & \ddots & \ddots & \ddots & \\ 0 & & \ddots & 0 & \sqrt{b_n} \\ & & & -\sqrt{b_n} & 0 \end{bmatrix},$$

$$b_i\in\mathbb{R}, b_i\neq 0,\ i=1,\cdots,n \qquad (7.11)$$

称为 **Routh 矩阵**.

Routh 矩阵会在线性控制理论中遇到.

6.7.8 定理 设 $A\in\mathbb{R}^{n\times n}$ 是由(7.10)确定的 Schwarz 矩阵,则

(1) A 为稳定矩阵(即其所有特征值具有负实部)的确充分必要条件是 $b_i>0, i=1,\cdots,n$.

(2) 由(7.11)确定的 Routh 矩阵 R 相似于 A.

证 (1) 证明从略.

(2) 取

$$Y=E\mathrm{diag}\left((-1)^{n+1}\sqrt{b_1\cdots b_n},(-1)^n\sqrt{b_1\cdots b_{n-1}},\cdots,-\sqrt{b_1b_2},\sqrt{b_1}\right), \qquad (7.12)$$

其中 $E=\begin{bmatrix}e_1 & \cdots & e_n\end{bmatrix}$ 是反序单位矩阵,则有

$$YAY^{-1}=R. \qquad \square$$

6.7.9 定理 如果(7.1)中的同伴矩阵 A 具有 n 个不同的特征值 $\lambda_1,\cdots,\lambda_n$,那么相应的特征向量是

$$u(\lambda_j)=\left(1,p_1(\lambda_j),\cdots,p_{n-1}(\lambda_j)\right)^{\mathrm{T}},\ j=1,\cdots,n, \qquad (7.13)$$

而且

$$M^{-1}AM=\mathrm{diag}(\lambda_1,\cdots,\lambda_n). \qquad (7.14)$$

这里,$p_0(\lambda),\cdots,p_{n-1}(\lambda)$是由(7.4)产生的,

$$M=\begin{bmatrix}u(\lambda_1) & \cdots & u(\lambda_n)\end{bmatrix}. \qquad (7.15)$$

证明留作练习.

注意,当 A 简化为友矩阵时,(7.15)中的矩阵 M 便简化成为 Vandermonde 矩阵.

6.8 结式矩阵

6.8.1 引言 稳定性问题的进展有赖两个广阔领域的研究.第一是矩阵惯性理论的研究,见 **3.14**;第二是关于结式矩阵性质的研究.

先看一种简单的特殊情形.可以证明:二次方程

$$a_0\lambda^2+a_1\lambda+a_2=0\ (a_0\neq 0)$$

和

$$b_0\lambda^2+b_1\lambda+b_2=0$$

有一个公根的充分必要条件是

$$\begin{vmatrix} a_0 & a_1 & a_2 & 0 \\ 0 & a_0 & a_1 & a_2 \\ b_0 & b_1 & b_2 & 0 \\ 0 & b_0 & b_1 & b_2 \end{vmatrix}=0. \qquad (8.1)$$

(8.1)中的行列式是两个多项式

$$a(\lambda)=a_0\lambda^2+a_1\lambda+a_2$$

和

$$b(\lambda)=b_0\lambda^2+b_1\lambda+b_2$$

的一个结式.

一般,设

$$a(\lambda)=a_0\lambda^n+a_1\lambda^{n-1}+\cdots+a_{n-1}\lambda+a_n, a_0\neq 0 \quad (8.2)$$

和

$$b(\lambda)=b_0\lambda^m+b_1\lambda^{m-1}+\cdots+b_{m-1}\lambda+b_m, b_0\neq 0 \quad (8.3)$$

是两个任意的复系数多项式,$n\geq m$.相伴多项式$a(\lambda)$和$b(\lambda)$的**结式**(resultant)是指关于系数

$$a_0,\cdots,a_n$$

和

$$b_0,\cdots,b_m$$

的纯量函数,其值等于零当且仅当$a(\lambda)$和$b(\lambda)$互素(亦即没有次数大于零的公因式或说没有共同零点或公根).

6.8.2 定义 设$a(\lambda)$和$b(\lambda)$是形如(8.2)和(8.3)的多项式,方阵A称为相伴$a(\lambda)$和$b(\lambda)$的**结式矩阵**(resultant matrix),如果$\det A$是相伴$a(\lambda)$和$b(\lambda)$的结式.

6.8.3 定理 设$a(\lambda)$和$b(\lambda)$是形如(8.2)和(8.3)的多项式,$a_0=1$,$n\geq m$,又设

$$C=\begin{bmatrix} 0 & 1 & & & \\ & \ddots & \ddots & 0 & \\ 0 & & \ddots & \ddots & \\ & & & 0 & 1 \\ -a_n & -a_{n-1} & \cdots & -a_2 & -a_1 \end{bmatrix} \quad (8.4)$$

是$a(\lambda)$的友矩阵,则矩阵

$$b(C)=b_0C^m+b_1C^{m-1}+\cdots+b_{m-1}C+b_mI \quad (8.5)$$

是相伴$a(\lambda)$和$b(\lambda)$的结式矩阵.

证　显然,$\det b(C)$是$a_0,\cdots,a_n$和$b_0,\cdots,b_m$的纯量函数.另一方面,根据(8.5),$b(C)$的特征值是

$$b(\lambda_1),\cdots,b(\lambda_n),\quad \lambda_1,\cdots,\lambda_n\in\lambda(C),$$

因此

$$\det b(C)=b(\lambda_1)\cdots b(\lambda_n),$$

从而,$b(C)$非奇异的充分必要条件是

$$b(\lambda_i)\neq 0,\ i=1,\cdots,n. \tag{8.6}$$

由于$a(\lambda)$就是矩阵C的特征多项式,$a(\lambda)$的根就是C的特征值,因而,条件(8.6)等价于$a(\lambda)$和$b(\lambda)$互质. □

现在推导矩阵$b(C)$的形式.设$b(C)$的行向量为

$$r^{(1)},\cdots,r^{(n)}.$$

注意到C的各行为

$$e_i^{\mathrm T}C=e_{i+1}^{\mathrm T},\ i=1,\cdots,n-1,$$
$$e_n^{\mathrm T}C=(-a_n,\cdots,-a_1),$$

推出$C^2,\cdots,C^{m-1}$的第1行为

$$\begin{aligned}e_1^{\mathrm T}C^k&=\left(e_1^{\mathrm T}C\right)C^{k-1}\\&=e_2^{\mathrm T}C^{k-1}=\cdots=e_k^{\mathrm T}C=e_{k+1}^{\mathrm T},\\&\qquad\qquad k=1,\cdots,m-1;\end{aligned}$$

C^m的第1行为

$$e_1^{\mathrm T}C^m=e_m^{\mathrm T}C=\begin{cases}e_{m+1}^{\mathrm T}, & m<n,\\(-a_n,\cdots,-a_1), & m=n.\end{cases}$$

于是,$b(C)$的第1行为

$$\begin{aligned}r^{(1)}&=e_1^{\mathrm T}b(C)\\&=b_0e_1^{\mathrm T}C^m+b_1e_1^{\mathrm T}C^{m-1}+\cdots+b_{m-1}e_1^{\mathrm T}C+b_me_1^{\mathrm T}I\\&=b_0e_1^{\mathrm T}C^m+b_1e_m^{\mathrm T}+\cdots+b_{m-1}e_2^{\mathrm T}+b_me_1^{\mathrm T}\end{aligned}$$

$$=\begin{cases}(b_m,b_{m-1},\cdots,b_1,b_0,0,\cdots,0), & m<n,\\ (b_n-b_0a_n,\cdots,b_1-b_0a_1), & m=n;\end{cases} \tag{8.7}$$

$b(C)$的第2至n行为

$$\begin{aligned}r^{(i)}&=e_i^{\mathrm{T}}b(C)\\ &=e_{i-1}^{\mathrm{T}}Cb(C)=e_{i-1}^{\mathrm{T}}b(C)C=r^{(i-1)}C\\ &=\cdots=r^{(1)}C^{i-1}, \qquad i=2,\cdots,n;\end{aligned}$$

从而矩阵$b(C)$有表示式

$$b(C)\equiv\begin{bmatrix}r^{(1)}\\ r^{(2)}\\ r^{(3)}\\ \vdots\\ r^{(n)}\end{bmatrix}=\begin{bmatrix}r^{(1)}\\ r^{(1)}C\\ r^{(1)}C^2\\ \vdots\\ r^{(1)}C^{n-1}\end{bmatrix}. \tag{8.8}$$

6.8.4 定义 设多项式

$$a(\lambda)=\sum_{i=0}^{n}a_i\lambda^{n-i},a_0\neq 0;\quad b(\lambda)=\sum_{i=0}^{n}b_i\lambda^{n-i}, \tag{8.9}$$

这蕴涵着$a(\lambda)$是n次多项式,$b(\lambda)$是次数不超过n(可以低于n)的多项式.表示式

$$\frac{a(\lambda)b(\mu)-a(\mu)b(\lambda)}{\lambda-\mu}$$

显然是λ和μ的二元多项式.记

$$\sum_{i=1}^{n}\sum_{j=1}^{n}z_{ij}\lambda^{i-1}\mu^{j-1}\equiv\frac{a(\lambda)b(\mu)-a(\mu)b(\lambda)}{\lambda-\mu} \tag{8.10}$$

和

$$Z\equiv[z_{ij}], \tag{8.11}$$

Z是$n\times n$的对称矩阵,称其为相伴多项式$a(\lambda)$和$b(\lambda)$的 **Bézout**

矩阵,也记作 $\mathrm{Bez}(a,b)$.

显然,

$$\mathrm{Bez}(a,b) = -\mathrm{Bez}(b,a). \tag{8.12}$$

一般地说,Bézout 矩阵是对称矩阵的一种特殊情形,但是不能凭直观确定一个具体的对称矩阵是不是 Bézout 矩阵.

例 对于 $n=3$,从定义出发,展开(8.10)右端,然后根据(8.11)得

$$Z \equiv \mathrm{Bez}(a,b)$$

$$= \begin{bmatrix} a_2b_3 - b_2a_3 & a_1b_3 - b_1a_3 & a_0b_3 - b_0a_3 \\ a_1b_3 - b_1a_3 & (a_0b_3 - b_0a_3) + (a_1b_2 - b_1a_2) & a_0b_2 - b_0a_2 \\ a_0b_3 - b_0a_3 & a_0b_2 - b_0a_2 & a_0b_1 - b_0a_1 \end{bmatrix}.$$

从定义出发可以得到 $\mathrm{Bez}(a,b)$ 的显式表示,然而实现起来并不容易,建立 $\mathrm{Bez}(a,b)$ 显式表示的捷径是和证明其为结式矩阵联系在一起的.

6.8.5 定义 设多项式

$$a(\lambda) = \sum_{i=0}^{n} a_i \lambda^{n-i}, a_0 \neq 0, \tag{8.13}$$

由 $a(\lambda)$ 的系数确定的三角 Hankel 矩阵

$$S(a) = \begin{bmatrix} a_{n-1} & a_{n-2} & \cdots & a_0 \\ a_{n-2} & & \iddots & 0 \\ \vdots & \iddots & \iddots & \vdots \\ a_0 & 0 & \cdots & 0 \end{bmatrix} \tag{8.14}$$

称为 $a(\lambda)$ 的**对称化子**(symmetrizer).

注意,对称化子 $S(a)$ 不依赖 $a(\lambda)$ 的尾项系数 a_n.

现令多项式

$$\hat{a}(\lambda) \equiv \lambda^n a(\lambda^{-1}) = a_n\lambda^n + a_{n-1}\lambda^{n-1} + \cdots + a_1\lambda + a_0, \quad (8.15)$$

则其对称化子为

$$S(\hat{a}) = \begin{bmatrix} a_1 & a_2 & \cdots & a_n \\ a_2 & & \unicode{x22F0} & 0 \\ \vdots & \unicode{x22F0} & \unicode{x22F0} & \vdots \\ a_n & 0 & \cdots & 0 \end{bmatrix}. \quad (8.16)$$

利用反序单位矩阵 $E = [e_n \quad \cdots \quad e_1]$,得 $ES(a)$ 和 $S(a)E$ 是 Toeplitz 矩阵:

$$ES(a) = \begin{bmatrix} a_0 & 0 & \cdots & 0 \\ \vdots & \ddots & \ddots & \vdots \\ a_{n-2} & & \ddots & 0 \\ a_{n-1} & a_{n-2} & \cdots & a_0 \end{bmatrix} \quad (8.17)$$

和

$$S(a)E = \begin{bmatrix} a_0 & \cdots & a_{n-2} & a_{n-1} \\ 0 & \ddots & & a_{n-2} \\ \vdots & \ddots & \ddots & \vdots \\ 0 & \cdots & 0 & a_0 \end{bmatrix}. \quad (8.18)$$

对称化子是经常出现的矩阵,从下述性质便可以知其取名的理由.

6.8.6 引理 设 $a(\lambda)$ 是形如(8.13)的多项式,$S(a)$ 是形如(8.14)的 $a(\lambda)$ 的对称化子,C_a 是 $a(\lambda)$ 的友矩阵

$$C_a=\begin{bmatrix} 0 & 1 & & & \\ & \ddots & \ddots & & 0 \\ & 0 & & 0 & 1 \\ -\frac{a_n}{a_0} & \cdots & -\frac{a_2}{a_0} & & -\frac{a_1}{a_0} \end{bmatrix},$$

则 $S(a)C_a$ 是对称矩阵,

$$S(a)C_a=\begin{bmatrix} -a_n & 0 & 0 & \cdots & 0 \\ 0 & a_{n-2} & a_{n-3} & \cdots & a_0 \\ 0 & a_{n-3} & & \iddots & 0 \\ \vdots & \vdots & \iddots & \iddots & \vdots \\ 0 & a_0 & 0 & \cdots & 0 \end{bmatrix}.$$

而且,对于 $1\le k\le n-1$,成立

$$S(a)C_a^k=\operatorname{diag}(-ES_1E,S_2),\tag{8.19}$$

其中 $E=[e_k \quad \cdots \quad e_1]$; S_1 为 $k\times k$ 矩阵,是多项式

$$a_1(\lambda)=a_n\lambda^{n-k}+\cdots+a_{k+1}\lambda+a_k$$

的对称化子; S_2 为 $(n-k)\times(n-k)$ 矩阵,是多项式

$$a_2(\lambda)=a_0\lambda^k+\cdots+a_{k-1}\lambda+a_k$$

的对称化子.特别,

$$S(a)C_a^n=-ES(\hat{a})E,\tag{8.20}$$

$S(\hat{a})$ 是形如(8.16)的 $\hat{a}(\lambda)$ 的对称化子.

(8.19)和(8.20)表明, $S(a)C_a^k$ $(1\le k\le n)$ 都是对称矩阵.

证明只须直接计算,留作练习.

6.8.7 定理 设 $a(\lambda)$ 和 $b(\lambda)$ 是形如(8.9)的多项式; $S(a)$, $S(b)$ 和 $S(\hat{a})$, $S(\hat{b})$ 是 **6.8.5** 定义的对称化子,则

$$\mathrm{Bez}(a,b)=S(a)S(\hat{b})E-S(b)S(\hat{a})E. \tag{8.21}$$

证 将(8.10)写成

$$(\lambda-\mu)\sum_{i=1}^{n}\gamma_i(\mu)\lambda^{i-1}=a(\lambda)b(\mu)-a(\mu)b(\lambda), \tag{8.22}$$

其中

$$\gamma_i(\mu)=\sum_{j=1}^{n}z_{ij}\mu^{j-1}, \quad i=1,\cdots,n \tag{8.23}$$

是多项式.比较(8.22)等号两端λ^{i-1}的系数,得

$$\gamma_i(\mu)-\mu\gamma_{i+1}(\mu)=a_{n-i}b(\mu)-b_{n-i}a(\mu),$$
$$i=1,\cdots,n-1;$$
$$\gamma_n(\mu)=a_0b(\mu)-b_0a(\mu).$$

由此从后向前递推,有

$$\gamma_i(\mu)=\left(a_0\mu^{n-i}+a_1\mu^{n-i-1}+\cdots+a_{n-i-1}\mu+a_{n-i}\right)b(\mu)$$
$$-\left(b_0\mu^{n-i}+b_1\mu^{n-i-1}+\cdots+b_{n-i-1}\mu+b_{n-i}\right)a(\mu),$$
$$i=n,n-1,\cdots,1, \tag{8.24}$$

因(8.23), $\gamma_i(\mu)$的次数$\le n-1$, $i=1,\cdots,n$.令$\mathrm{Bez}(a,b)$的行为

$$z^{(i)}=(z_{i1},\cdots,z_{in})^{\mathrm{T}}, \quad i=1,\cdots,n.$$

从(8.23)和(8.24)推出

$$\left(z^{(i)}\right)^{\mathrm{T}}\begin{pmatrix}1\\ \mu\\ \vdots\\ \mu^{n-1}\end{pmatrix}=(a_{n-i},\cdots,a_1,a_0,0,\cdots,0)\begin{pmatrix}b(\mu)\\ b(\mu)\mu\\ \vdots\\ b(\mu)\mu^{n-1}\end{pmatrix}$$
$$-(b_{n-i},\cdots,b_1,b_0,0,\cdots,0)\begin{pmatrix}a(\mu)\\ a(\mu)\mu\\ \vdots\\ a(\mu)\mu^{n-1}\end{pmatrix}. \tag{8.25}$$

容易验证

$$\begin{pmatrix} a(\mu) \\ a(\mu)\mu \\ \vdots \\ a(\mu)\mu^{n-1} \end{pmatrix} = \begin{bmatrix} a_n & a_{n-1} & \cdots & a_1 \\ 0 & \ddots & \ddots & \vdots \\ \vdots & \ddots & \ddots & a_{n-1} \\ 0 & \cdots & 0 & a_n \end{bmatrix} \begin{pmatrix} 1 \\ \mu \\ \vdots \\ \mu^{n-1} \end{pmatrix}$$

$$+ \begin{bmatrix} a_0 & 0 & \cdots & 0 \\ a_1 & \ddots & \ddots & \vdots \\ \vdots & \ddots & \ddots & 0 \\ a_{n-1} & \cdots & a_1 & a_0 \end{bmatrix} \begin{pmatrix} 1 \\ \mu \\ \vdots \\ \mu^{n-1} \end{pmatrix} \mu^n$$

$$= S(\hat{a})E \begin{pmatrix} 1 \\ \mu \\ \vdots \\ \mu^{n-1} \end{pmatrix} + ES(a) \begin{pmatrix} 1 \\ \mu \\ \vdots \\ \mu^{n-1} \end{pmatrix} \mu^n .$$

自然,对 $b(\mu)$ 成立同样的关系式.于是,由(8.25)有

$$\operatorname{Bez}(a,b) \begin{pmatrix} 1 \\ \mu \\ \vdots \\ \mu^{n-1} \end{pmatrix} = S(a)S(\hat{b})E \begin{pmatrix} 1 \\ \mu \\ \vdots \\ \mu^{n-1} \end{pmatrix} + S(a)ES(b) \begin{pmatrix} 1 \\ \mu \\ \vdots \\ \mu^{n-1} \end{pmatrix} \mu^n$$

$$- S(b)S(\hat{a})E \begin{pmatrix} 1 \\ \mu \\ \vdots \\ \mu^{n-1} \end{pmatrix} - S(b)ES(a) \begin{pmatrix} 1 \\ \mu \\ \vdots \\ \mu^{n-1} \end{pmatrix} \mu^n . \quad (8.26)$$

注意到对任何 $a(\lambda)$ 和 $b(\lambda)$,

$$S(a)ES(b) = S(b)ES(a), S(\hat{a})ES(\hat{b}) = S(\hat{b})ES(\hat{a}), \quad (8.27)$$

(8.26)成为

$$\text{Bez}(a,b)\begin{pmatrix}1\\ \mu\\ \vdots\\ \mu^{n-1}\end{pmatrix}=\left[S(a)S(\hat{b})E-S(b)S(\hat{a})E\right]\begin{pmatrix}1\\ \mu\\ \vdots\\ \mu^{n-1}\end{pmatrix}.$$

由于 $\mu\in\mathbb{C}$是任意的,因此对不同的 $\mu_1,\cdots\mu_n\in\mathbb{C}$成立

$$\begin{aligned}&\text{Bez}(a,b)V(\mu_1,\cdots,\mu_n)\\&\qquad=\left[S(a)S(\hat{b})E-S(b)S(\hat{a})E\right]V(\mu_1,\cdots,\mu_n),\end{aligned}$$

其中 $V(\mu_1,\cdots,\mu_n)$是非奇异的 Vandermonde 矩阵(见 **2.6.1**),此等式蕴涵(8.21). □

6.8.8 定理 设 $a(\lambda)$ 和 $b(\lambda)$ 是形如(8.9)的多项式;C_a 和 C_b 是如同 **6.8.6** 中的形式的友矩阵,$S(a)$ 和 $S(b)$ 是 **6.8.5** 定义的对称化子,则

$$\text{Bez}(a,b)=S(a)b(C_a)=-S(b)a(C_b). \tag{8.28}$$

证 对于任意的次数不超过 n 的多项式 $b_1(\lambda)$ 和 $b_2(\lambda)$ 及任意的 $\alpha,\beta\in\mathbb{C}$,显然有

$$\text{Bez}(a,\alpha b_1+\beta b_2)=\alpha\text{Bez}(a,b_1)+\beta\text{Bez}(a,b_2),$$

这表明当多项式 a 固定时,矩阵 $\text{Bez}(a,b)$ 关于多项式 b 是线性的.因此,对于 $b(\lambda)=\sum_{i=0}^{n}b_i\lambda^{n-i}$,有

$$\text{Bez}(a,b)=\sum_{i=0}^{n}b_i\text{Bez}(a,\lambda^{n-i}).$$

注意到根据(8.10)和(8.11),$\text{Bez}(a,\lambda^k)$是展式

$$\begin{aligned}\frac{a(\lambda)\mu^k-\lambda^k a(\mu)}{\lambda-\mu}&=\sum_{i=0}^{n}a_i\frac{\lambda^i\mu^k-\lambda^k\mu^i}{\lambda-\mu}\\&=-\sum_{i=0}^{k-1}a_i\left(\lambda^{k-1}\mu^i+\cdots+\lambda^i\mu^{k-1}\right)\end{aligned}$$

$$+\sum_{i=k+1}^{n} a_i\left(\lambda^{i-1}\mu^k+\cdots+\lambda^k\mu^{i-1}\right)$$

的系数确定的矩阵,检验(8.19)中矩阵的形式,可以发现

$$\operatorname{Bez}\left(a,\lambda^k\right)=\operatorname{diag}\left(-ES_1E,S_2\right)=S(a)C_a^k,$$

由此即得

$$\operatorname{Bez}(a,b)=S(a)b\left(C_a\right),$$

(8.28)的第一个等号成立;再利用(8.12),推出(8.28)的第二个等号也成立. □

(8.28)称为矩阵 $\operatorname{Bez}(a,b)$ 的 **Barnett 分解**.

现在推导 Barnett 分解的另一表示形式.依 Cayley-Hamilton 定理 **1.6.9**,

$$a\left(C_a\right)=0,$$

因此

$$b\left(C_a\right)=b\left(C_a\right)-\frac{b_0}{a_0}a\left(C_a\right)=\sum_{i=1}^{n}\tilde{b}_iC_a^{n-i},$$

其中

$$\tilde{b}_i=b_i-\frac{b_0}{a_0}a_i,\quad i=1,\cdots,n.$$

定义

$$\eta=\left(\tilde{b}_n,\cdots,\tilde{b}_1\right)^{\mathrm{T}}.$$

因为

$$e_1^{\mathrm{T}}C_a^j=\left(e_1^{\mathrm{T}}C_a\right)C_a^{j-1}=e_2^{\mathrm{T}}C_a^{j-1}=e_{j+1}^{\mathrm{T}},\qquad j=1,\cdots,n-1,\tag{8.29}$$

得 $b\left(C_a\right)$ 的第1行为

$$e_1^{\mathrm{T}}b\left(C_a\right)=\sum_{i=1}^{n}\tilde{b}_ie_1^{\mathrm{T}}C_a^{n-i}=\sum_{i=1}^{n}\tilde{b}_ie_{n-i+1}^{\mathrm{T}}=\left(\tilde{b}_n,\cdots,\tilde{b}_1\right)=\eta^{\mathrm{T}};$$

由此及(8.29)，$b(C_a)$的第 2 至 n 行为

$$e_j^{\mathrm{T}}b(C_a)=e_1^{\mathrm{T}}C_a^{j-1}b(C_a)=e_1^{\mathrm{T}}b(C_a)C_a^{j-1}=\eta^{\mathrm{T}}C_a^{j-1},\quad j=2,\cdots,n.$$

因此

$$b(C_a)=\begin{bmatrix}\eta^{\mathrm{T}}\\ \eta^{\mathrm{T}}C_a\\ \vdots\\ \eta^{\mathrm{T}}C_a^{n-1}\end{bmatrix}.$$

这样，连同(8.28)，得分解式

$$\mathrm{Bez}(a,b)=S(a)\begin{bmatrix}\eta^{\mathrm{T}}\\ \eta^{\mathrm{T}}C_a\\ \vdots\\ \eta^{\mathrm{T}}C_a^{n-1}\end{bmatrix}. \tag{8.30}$$

6.8.9 定理 设 $a(\lambda)$ 和 $b(\lambda)$ 是形如(8.9)的多项式. $a(\lambda)$ 和 $b(\lambda)$ 没有共同零点的充分必要条件是相伴的 $\mathrm{Bez}(a,b)$ 为非奇异矩阵.

证 根据(8.28)，

$$\mathrm{Bez}(a,b)=S(a)b(C_a).$$

由于 $\det S(a)=a_0^n\neq 0$，从而，$\det(\mathrm{Bez}(\mathrm{a},\mathrm{b}))\neq 0$ 的充分必要条件是 $\det b(C_a)\neq 0$. 因为 $a(\lambda)$ 是友矩阵 C_a 的特征多项式，其零点 $\lambda_1,\cdots,\lambda_n$ 即 C_a 的特征值，而 $b(C_a)$ 的特征值是 $b(\lambda_1),\cdots,b(\lambda_n)$，因此

$\mathrm{Bez}(a,b)$ 非奇异

$\Leftrightarrow \det b(C_a)\neq 0 \Leftrightarrow b(\lambda_i)\neq 0,\ i=1,\cdots,n$

$\Leftrightarrow a(\lambda)$ 和 $b(\lambda)$ 没有共同零点. □

这个定理表明了 $\mathrm{Bez}(a,b)$ 是相伴多项式 $a(\lambda)$ 和 $b(\lambda)$ 的结式

矩阵.

6.8.10 定义 设多项式

$$a(\lambda)=\sum_{i=0}^{n}a_i\lambda^{n-i},a_0\neq 0;\quad b(\lambda)=\sum_{i=0}^{m}b_i\lambda^{m-i},b_0\neq 0,$$

而且 $n\geq m$. $(m+n)\times(m+n)$ 矩阵

$$R(a,b)=\begin{bmatrix} a_0 & a_1 & \cdots & \cdots & a_n & & 0 \\ 0 & \ddots & \ddots & & & \ddots & \\ & & a_0 & a_1 & \cdots & \cdots & a_n \\ b_0 & b_1 & \cdots & b_m & & & \\ & \ddots & \ddots & & \ddots & & 0 \\ & 0 & \ddots & \ddots & & \ddots & \\ & & & b_0 & b_1 & \cdots & b_m \end{bmatrix} \tag{8.31}$$

称为相伴多项式 $a(\lambda)$ 和 $b(\lambda)$ 的 **Sylvester 矩阵**.

下述定理表明 Sylvester 矩阵 $R(a,b)$ 是相伴多项式 $a(\lambda)$ 和 $b(\lambda)$ 的结式矩阵.

6.8.11 定理 设多项式

$$a(\lambda)=\sum_{i=0}^{n}a_i\lambda^{n-i},a_0\neq 0;\quad b(\lambda)=\sum_{i=0}^{m}b_i\lambda^{m-i},b_0\neq 0,$$

而且 $n\geq m$. $a(\lambda)$ 和 $b(\lambda)$ 没有共同零点的充分必要条件是相伴的 $R(a,b)$ 为非奇异矩阵.

证 视 $b(\lambda)$ 为多项式

$$q(\lambda)=\sum_{i=0}^{n}q_i\lambda^{n-i},$$

其中

$$q_0=\cdots=q_{n-m-1}=0;\quad q_{n-m+k}=b_k,k=0,1,\cdots,m.$$

考虑 $2n\times 2n$ 矩阵

$$R(a,q)=\begin{bmatrix} a_0 & a_1 & \cdots & a_n & & 0 \\ 0 & \ddots & \ddots & & \ddots & \\ & & a_0 & a_1 & \cdots & a_n \\ q_0 & q_1 & \cdots & q_n & & 0 \\ & \ddots & \ddots & & \ddots & \\ 0 & & q_0 & q_1 & \cdots & q_n \end{bmatrix}. \tag{8.32}$$

一方面,利用 **6.8.5** 中的记号,可将 $R(a,q)$ 写成

$$R(a,q)=\begin{bmatrix} S(a)E & ES(\hat{a}) \\ S(q)E & ES(\hat{q}) \end{bmatrix},$$

应用(8.21)和(8.27),推出

$$\begin{bmatrix} I & 0 \\ -S(q)E & S(a)E \end{bmatrix} R(a,q)=\begin{bmatrix} S(a)E & ES(\hat{a}) \\ 0 & \mathrm{Bez}(a,q)E \end{bmatrix},$$

由此,及 $S(a)$ 非奇异,得知 $R(a,q)$ 非奇异的充分必要条件是 $\mathrm{Bez}(a,q)$ 非奇异.

另一方面,(8.32)又有分块形式

$$R(a,q)=\left[\begin{array}{cc|cccccc} a_0 & \cdots & a_{n-m} & \cdots & a_n & & & 0 \\ & \ddots & & \ddots & & \ddots & & \\ & & \ddots & & \ddots & & \ddots & \\ & & & a_0 & \cdots & a_{n-m} & \cdots & a_n \\ \hline & & b_0 & \cdots & b_m & & & 0 \\ & & & \ddots & & \ddots & & \\ & 0 & & & \ddots & & \ddots & \\ & & & & & b_0 & \cdots & b_m \end{array}\right]$$

$$
=\begin{bmatrix} T & * \\ 0 & R(a,b) \end{bmatrix}, \quad T=\begin{bmatrix} a_0 & \cdots & a_{n-m-1} \\ & \ddots & \vdots \\ 0 & & a_0 \end{bmatrix}.
$$

由此,及 $a_0 \neq 0$,得知 $R(a,b)$ 非奇异当且仅当 $R(a,q)$ 非奇异.

最后,综合以上两个方面,并利用 **6.8.9**,得

$R(a,b)$ 非奇异 $\Leftrightarrow$ $R(a,q)$ 非奇异 $\Leftrightarrow$ $\mathrm{Bez}(a,q)$ 非奇异

$\Leftrightarrow$ $a(\lambda)$ 和 $q(\lambda)$ 即 $b(\lambda)$ 没有共同零点. □

6.9 Hurwitz 矩阵和 Schur-Cohn 矩阵

6.9.1 定义 设多项式

$$
a(\lambda)=\sum_{i=0}^{n} a_i \lambda^{n-i}, \quad a_0 \neq 0,
$$

用 $i_+(a), i_-(a)$ 和 $i_0(a)$ 分别表示 $a(\lambda)$ 按重数计的具有正实部,负实部和零实部的零点的个数,

$$
i_+(a)+i_-(a)+i_0(a)=n,
$$

三元组

$$
i(a)=(i_+(a), i_-(a), i_0(a)) \tag{9.1}
$$

称为 $a(\lambda)$ **关于虚轴的惯性**.

特别,当

$$
i_+(a)=i_0(a)=0
$$

时,即 $a(\lambda)$ 的所有零点具有负实部,称 $a(\lambda)$ 为关于虚轴的**稳定多项式**.

相对地,三元组

$$
i'(a)=(i'_+(a), i'_-(a), i'_0(a)), \tag{9.2}
$$

其中 $i'_+(a), i'_-(a)$ 和 $i'_0(a)$ 分别表示 $a(\lambda)$ 按重数计的位于开上半平面,开下半平面和实轴的零点的个数, $i'(a)$ 称为 $a(\lambda)$ **关于实轴的惯性**.

显然,若 $\tilde{a}(\lambda) \equiv a(-i\lambda)$,则

$$i'(\tilde{a}) = i(a). \tag{9.3}$$

当 $a(\lambda)$ 是实系数多项式时,其复零点共轭成对出现,因此

$$i'_+(a) = i'_-(a). \tag{9.4}$$

在稳定性问题中,稳定多项式的讨论非常重要.一种途径是通过构造相伴多项式的矩阵,建立检测多项式所有零点位于开左半复平面的充分必要条件.在实际应用中,通常只须考虑实系数多项式.

6.9.2 定义 设多项式 $a(\lambda) = \sum_{i=0}^{n} a_i \lambda^{n-i}, a_0 \neq 0$; $n \times n$ 矩阵

$$H = \begin{bmatrix} a_1 & a_3 & a_5 & \cdots & a_{2n-3} & a_{2n-1} \\ a_0 & a_2 & a_4 & \cdots & a_{2n-4} & a_{2n-2} \\ 0 & a_1 & a_3 & \cdots & a_{2n-5} & a_{2n-3} \\ 0 & a_0 & a_2 & \cdots & a_{2n-6} & a_{2n-4} \\ \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \\ 0 & 0 & 0 & \cdots & a_{n-2} & a_n \end{bmatrix}, \tag{9.5}$$

其中 $a_i = 0, \forall i > n$, H 称为相伴多项式 $a(\lambda)$ 的 **Hurwitz 矩阵**.

如果将形如(9.5)的 Hurwitz 矩阵 H 中的行重新排列,先第 1,3,5,⋯行,后第 2,4,6,⋯行,那么,当 $n = 2m$ 时, H 便成为相伴多项式

$$a_1\lambda^m + a_3\lambda^{m-1} + \cdots + a_{2m-1}\lambda$$

和

$$a_0\lambda^m + a_2\lambda^{m-1} + \cdots + a_{2m}$$

的 Sylvester 结式矩阵;当 $n = 2m+1$ 时,成为相伴多项式

$$a_1\lambda^m + a_3\lambda^{m-1} + \cdots + a_{2m+1}$$

和

$$a_0\lambda^{m+1} + a_2\lambda^m + \cdots + a_{2m}\lambda$$

的 Sylvester 结式矩阵.

6.9.3 Routh-Hurwitz 定理 设 $a(\lambda)$ 是 n 次实系数多项式,首项系数 $a_0 \neq 0$,相伴 $a(\lambda)$ 的 Hurwitz 矩阵的前主子式序列

$$\Delta_1, \Delta_2, \cdots, \Delta_n$$

全不为零.则 $a(\lambda)$ 没有纯虚数零点,而且

$$i_+(a) = \mathrm{V}\left(a_0, \Delta_1, \frac{\Delta_2}{\Delta_1}, \cdots, \frac{\Delta_n}{\Delta_{n-1}}\right),$$

$$i_-(a) = n - \mathrm{V}\left(a_0, \Delta_1, \frac{\Delta_2}{\Delta_1}, \cdots, \frac{\Delta_n}{\Delta_{n-1}}\right)$$

$$\equiv \mathrm{P}\left(a_0, \Delta_1, \frac{\Delta_2}{\Delta_1}, \cdots, \frac{\Delta_n}{\Delta_{n-1}}\right),$$

其中记号

$$\mathrm{V}(s_0, s_1, \cdots, s_n)$$

和

$$\mathrm{P}(s_0, s_1, \cdots, s_n)$$

分别表示实数序列 $s_0, s_1, \cdots, s_n$ 中正负号交替和不变的次数.

这是一个著名的稳定性定理.证明可以参见[LT1985]的 13.9 的定理1.

6.9.4 推论 设$a(\lambda)$是首项系数$a_0>0$的n次实系数多项式,$a(\lambda)$稳定的充分必要条件是

$$\Delta_1>0,\Delta_2>0,\cdots,\Delta_n>0. \tag{9.6}$$

证 应用 **6.9.3**.

$a(\lambda)$稳定,即$i_-(a)=n$,必有

$$\mathrm{V}\left(a_0,\Delta_1,\frac{\Delta_2}{\Delta_1},\cdots,\frac{\Delta_n}{\Delta_{n-1}}\right)=0, \tag{9.7}$$

这就是说,

$$a_0,\Delta_1,\frac{\Delta_2}{\Delta_1},\cdots,\frac{\Delta_n}{\Delta_{n-1}}$$

全为同号数.从而,由$a_0>0$推出(9.6)成立.

反之,(9.6)及$a_0>0$蕴涵(9.7).因此,$i_-(a)=n$. □

6.9.5 定义 设多项式

$$a(\lambda)=\sum_{i=0}^{n}a_i\lambda^{n-i},a_0\neq 0;$$

$2n\times 2n$矩阵

$$S=\begin{bmatrix}S_1 & S_2\\ S_2 & S_1\end{bmatrix}, \tag{9.8}$$

其中

$$S_1=\begin{bmatrix}a_0 & a_1 & \cdots & a_{n-1}\\ & \ddots & \ddots & \vdots\\ & & \ddots & a_1\\ 0 & & & a_0\end{bmatrix},S_2=\begin{bmatrix} & & & a_n\\ & 0 & \iddots & \vdots\\ & \iddots & & a_2\\ a_n & \cdots & a_2 & a_1\end{bmatrix}. \tag{9.9}$$

S称为相伴多项式$a(\lambda)$的 **Schur-Cohn 矩阵**.

例 $n=4$的 Schur-Cohn 矩阵形如

例 $n=4$ 的 Schur-Cohn 矩阵形如

$$S=\begin{bmatrix} a_0 & a_1 & a_2 & a_3 & & & & a_4 \\ & a_0 & a_1 & a_2 & & & a_4 & a_3 \\ & & a_0 & a_1 & & a_4 & a_3 & a_2 \\ & & & a_0 & a_4 & a_3 & a_2 & a_1 \\ & & & a_4 & a_0 & a_1 & a_2 & a_3 \\ & & a_4 & a_3 & & a_0 & a_1 & a_2 \\ & a_4 & a_3 & a_2 & & & a_0 & a_1 \\ a_4 & a_3 & a_2 & a_1 & & & & a_0 \end{bmatrix}. \tag{9.10}$$

注意,如果在(9.8)中,将后 n 列按相反顺序排列,然后将后 n 行按相反顺序排列,那么所得矩阵可以写成

$$\begin{bmatrix} S_1 & S_2E \\ ES_2 & ES_1E \end{bmatrix}, \tag{9.11}$$

这是相伴多项式 $a(\lambda)$ 及其**倒换多项式**(reverse polynomial)

$$\lambda^n a\left(\frac{1}{\lambda}\right)=\sum_{i=0}^{n} a_{n-i}\lambda^{n-i}$$

的 Sylvester 结式矩阵.

一般,任何 $n\times n$ 矩阵 S ,如(9.10)中每一虚线方框围成的主子矩阵称为矩阵 S 的**内矩阵**(inner matrix).

6.9.6 Schur-Cohn 定理 设 $a(\lambda)$ 是首项系数 $a_0\neq 0$ 的 n 次实系数多项式, $a(\lambda)$ 稳定的充分必要条件是其 Schur-Cohn 矩阵的所有内矩阵的行列式大于零.

证明从略.

7 特殊积矩阵和广义逆矩阵

7.1 Kronecker 积

7.1.1 定义 设 $A=[a_{ij}]\in\mathbb{C}^{m\times n}$, $B=[b_{ij}]\in\mathbb{C}^{p\times q}$.用表示 A 的每个元素 a_{ij} 和 B 数乘而成的 $mp\times nq$ 矩阵,其块形式如下:

$$A\otimes B=\begin{bmatrix} a_{11}B & a_{12}B & \cdots & a_{1n}B \\ a_{21}B & a_{22}B & \cdots & a_{2n}B \\ \cdots & \cdots & \cdots & \cdots \\ a_{m1}B & a_{m2}B & \cdots & a_{mn}B \end{bmatrix}, \qquad (1.1)$$

$A\otimes B$ 称为 A 和 B 的 **Kronecker 积**,也称为**直积**或**张量积**.

注意,一般

$$A\otimes B\neq B\otimes A,$$

即矩阵的 Kronecker 乘法不具交换律.

设 $x=(x_1,\cdots,x_n)^{\mathrm{T}}$, $y=(y_1,\cdots,y_n)^{\mathrm{T}}\in\mathbb{C}^n$.根据 Kronecker 积的定义,有

$$xy^{\mathrm{T}}=\begin{bmatrix} x_1y_1 & x_1y_2 & \cdots & x_1y_n \\ x_2y_1 & x_2y_2 & \cdots & x_2y_n \\ \cdots & \cdots & \cdots & \cdots \\ x_ny_1 & x_ny_2 & \cdots & x_ny_n \end{bmatrix}=\begin{bmatrix} x_1y^{\mathrm{T}} \\ x_2y^{\mathrm{T}} \\ \vdots \\ x_ny^{\mathrm{T}} \end{bmatrix}=x\otimes y^{\mathrm{T}}, \quad (1.2)$$

通常称 xy^{T} 为向量 x 与 y 的**外积**.

7.1.2 定义 设 $A\in\mathbb{C}^{m\times n}$. A 的第 k 次 Kronecker 幂 $A^{\otimes k}$ 递推地定义为

$$A^{\otimes 1}\equiv A;\quad A^{\otimes k}\equiv A\otimes A^{\otimes(k-1)},\ k=2,3,\cdots. \tag{1.3}$$

7.1.3 定理 Kronecker 积有如下基本性质:

(1) $(\alpha A)\otimes B=A\otimes(\alpha B)$,

$$\forall\alpha\in\mathbb{C},A\in\mathbb{C}^{m\times n},B\in\mathbb{C}^{p\times q}.$$

(2) $(A\otimes B)^*=A^*\otimes B^*$,

$$\forall A\in\mathbb{C}^{m\times n},B\in\mathbb{C}^{p\times q}.$$

(3) $(A\otimes B)\otimes C=A\otimes(B\otimes C)$,

$$\forall A\in\mathbb{C}^{m\times n},B\in\mathbb{C}^{p\times q},C\in\mathbb{C}^{r\times s}.$$

(4) $(A+B)\otimes C=A\otimes C+B\otimes C$,

$$\forall A,B\in\mathbb{C}^{m\times n},C\in\mathbb{C}^{p\times q}.$$

(5) $A\otimes(B+C)=A\otimes B+A\otimes C$,

$$\forall A\in\mathbb{C}^{m\times n},B,C\in\mathbb{C}^{p\times q}.$$

(6) 若 $A,B\in\mathbb{C}^{n\times n}$ 是自伴矩阵,则 $A\otimes B$ 也是自伴矩阵.

(7) $A\otimes B=0$ 的充分必要条件是 $A=0$ 或 $B=0$.

证明留作练习.

7.1.4 定义 矩阵 $A=[a_{ij}]\in\mathbb{C}^{m\times n}$ 的相伴向量

$$\mathrm{vec}A\in\mathbb{C}^{mn},$$

定义为 A 的元素按列的顺序排列而成的向量:

$$\mathrm{vec}A=(a_{11},\cdots,a_{m1},a_{12},\cdots,a_{m2},\cdots\cdots\cdots,a_{1n},\cdots,a_{mn})^{\mathrm{T}}. \tag{1.4}$$

容易证明(留作练习):

(1) 成立

$$\mathrm{tr}AB = \left(\mathrm{vec}A^{\mathrm{T}}\right)^{\mathrm{T}} \mathrm{vec}B, \quad \forall A \in \mathbb{C}^{m\times n}, B \in \mathbb{C}^{n\times m}. \tag{1.5}$$

(2) 定义

$$\langle B, A\rangle \equiv \mathrm{tr}A^{*}B, \quad \forall A, B \in \mathbb{C}^{m\times n}, \tag{1.6}$$

则$\langle \bullet,\bullet\rangle$是$\mathbb{C}^{m\times n}$上的一个内积.

7.1.5 引理 成立

$$(A\otimes B)(C\otimes D) = AC\otimes BD,$$

$$\forall A\in\mathbb{C}^{m\times n}, B\in\mathbb{C}^{p\times q}, C\in\mathbb{C}^{n\times k}, D\in\mathbb{C}^{q\times r}. \tag{1.7}$$

证 设$A=[a_{ih}], C=[c_{hj}]$,则有块形式

$$A\otimes B=\left[a_{ih}B\right], \quad C\otimes D=\left[c_{hj}D\right].$$

于是$(A\otimes B)(C\otimes D)$的(i,j)-块是

$$\left((A\otimes B)(C\otimes D)\right)_{ij} = \sum_{h=1}^{n} a_{ih}Bc_{hj}D = \left(\sum_{h=1}^{n} a_{ih}c_{hj}\right)BD,$$

这是AC的(i,j)-元素与BD的数乘,依 **7.1.1**,亦即$AC\otimes BD$的(i,j)-块. □

这是通常矩阵乘积与 Kronecker 乘积的混合积的一个基本性质.

7.1.6 推论 设$A\in\mathbb{C}^{n\times n}$和$B\in\mathbb{C}^{m\times m}$均为非奇异矩阵.则$A\otimes B$也是非奇异矩阵,而且

$$(A\otimes B)^{-1} = A^{-1}\otimes B^{-1}. \tag{1.8}$$

证 依 **7.1.5**,

$$\begin{aligned}(A\otimes B)\left(A^{-1}\otimes B^{-1}\right) &= AA^{-1}\otimes BB^{-1}\\ &= I_n\otimes I_m = I_{nm}.\end{aligned}$$

□

7.1.7 定理 设$A\in\mathbb{C}^{n\times n}$和$B\in\mathbb{C}^{m\times m}$,则

(1) 如果$\lambda \in \lambda(A)$且$x \in \mathbb{C}^n$是相应的A的特征向量，$\mu \in \lambda(B)$且$y \in \mathbb{C}^m$是相应的B的特征向量,则

$$\lambda\mu \in \lambda(A \otimes B).$$

并且$x \otimes y \in \mathbb{C}^{nm}$是相应的$A \otimes B$的特征向量.

(2) $A \otimes B$的每个特征值可表示为A与B的特征值的乘积.也就是说,如果

$$\lambda(A) = \{\lambda_1, \cdots, \lambda_n\}, \ \lambda(B) = \{\mu_1, \cdots, \mu_m\},$$

那么

$$\lambda(A \otimes B) = \left\{\lambda_i \mu_j : i = 1, \cdots, n; j = 1, \cdots m\right\}. \tag{1.9}$$

这里特征值均按代数重数计.

(3) 成立

$$\lambda(A \otimes B) = \lambda(B \otimes A). \tag{1.10}$$

证 根据假设,

$$Ax = \lambda x, \quad By = \mu y, \quad x \neq 0, \quad y \neq 0.$$

利用式(1.7),有

$$(A \otimes B)(x \otimes y) = Ax \otimes By = \lambda x \otimes \mu y = \lambda\mu(x \otimes y).$$

定理前一结论得证.下面证明余下的结论.

Schur 上三角标准形定理 **5.3.3** 保证存在酉矩阵$U \in \mathbb{C}^{n \times n}$和$V \in \mathbb{C}^{m \times m}$,使得

$$U^* A U = T_A$$

和

$$V^* B V = T_B,$$

其中T_A和T_B是上三角矩阵.由此并利用(1.7),得

$$(U \otimes V)^* (A \otimes B)(U \otimes V) = (U^* A U) \otimes (V^* B V) = T_A \otimes T_B,$$

表明$A \otimes B$相似于$T_A \otimes T_B$;显然,$T_A \otimes T_B$仍是上三角矩阵.因此,A,B和$A \otimes B$的特征值分别正好是T_A,T_B和$T_A \otimes T_B$的主对角元素,而$T_A \otimes T_B$的主对角元素由T_A的主对角元素和T_B的主对角元素的nm个两两之积组成. □

7.1.8 推论 设 $A \in \mathbb{C}^{n\times n}$ 和 $B \in \mathbb{C}^{m\times m}$ 是正定(半正定)矩阵.则 $A \otimes B$ 也是正定(半正定)矩阵.

证 由假设, A 和 B 是自伴的,依 **7.1.3** 的(6), $A \otimes B$ 也是自伴的;再依 **7.1.7**,如果 A 和 B 的特征值全为正的,那么 $A \otimes B$ 的特征值也全为正的,而特征值全为正的自伴矩阵是正定矩阵.

关于半正定的结论的推证是类似的. □

7.1.9 定理 设 $A \in \mathbb{C}^{m\times n}$ 和 $B \in \mathbb{C}^{p\times q}$ 有奇异值分解(见 **5.4.5**)

$$A = U_1^* \Sigma_1 V_1$$

和

$$B = U_2^* \Sigma_2 V_2,$$

其中

$$U_1 \in \mathbb{C}^{m\times m}, V_1 \in {}^{n\times n}, U_2 \in \mathbb{C}^{p\times p}, V_2 \in \mathbb{C}^{q\times q}$$

是酉矩阵,而且

$$\Sigma_1 = \begin{bmatrix} \Sigma_{r_1} & 0 \\ 0 & 0 \end{bmatrix} \in \mathbb{C}^{m\times n},\ \ \Sigma_{r_1} = \operatorname{diag}\left(\tilde{\sigma}_1, \cdots, \tilde{\sigma}_{r_1}\right),$$

$\tilde{\sigma}_1 \geq \cdots \geq \tilde{\sigma}_{r_1} > 0$ 是 A 的非零奇异值,和

$$\Sigma_2 = \begin{bmatrix} \Sigma_{r_2} & 0 \\ 0 & 0 \end{bmatrix} \in \mathbb{C}^{p\times q},\ \ \Sigma_{r_2} = \operatorname{diag}\left(\hat{\sigma}_1, \cdots, \hat{\sigma}_{r_2}\right),$$

$\hat{\sigma}_1 \geq \cdots \geq \hat{\sigma}_{r_2} > 0$ 是 B 的非零奇异值.则

$$A \otimes B = \left(U_1 \otimes U_2\right)^* \left(\Sigma_1 \otimes \Sigma_2\right)\left(V_1 \otimes V_2\right).$$

而且, $A \otimes B$ 的非零奇异值是 $r_1 r_2$ 个正数(按代数重数计):

$$\tilde{\sigma}_i \hat{\sigma}_j,\ \ i = 1, \cdots, r_1,\ \ j = 1, \cdots, r_2.$$

$A \otimes B$ 的零奇异值的重数为 $\min\{mp, nq\} - r_1 r_2$.

特别, $A \otimes B$ 和 $B \otimes A$ 有相同的奇异值,而且

$$\operatorname{rank}(A \otimes B) = \operatorname{rank}(B \otimes A) = r_1 r_2.$$

证 从混合积性质 **7.1.5** 直接推出.详细推证留作练习. □

7.1.10 引理　设 $A \in \mathbb{C}^{m\times n}$, $B \in \mathbb{C}^{p\times q}$ 和 $C \in \mathbb{C}^{m\times q}$ 是给定的，而 $X \in \mathbb{C}^{n\times p}$ 是未知的.则矩阵方程

$$AXB = C$$

等价于含 np 个未知数的具有 mq 个方程的方程组

$$\left(B^{\mathrm{T}} \otimes A\right)\mathrm{vec}X = \mathrm{vec}C,$$

亦即

$$\mathrm{vec}(AXB) = \left(B^{\mathrm{T}} \otimes A\right)\mathrm{vec}X.$$

证　对于任一矩阵 Q，用 Q_k 表示其第 k 列.设 $B = [b_{ij}]$.因为

$$(AXB)_k = A(XB)_k = AXB_k = A\left(\sum_{i=1}^{p} b_{ik} X_i\right)$$

$$= \left(b_{1k}A, b_{2k}A, \cdots, b_{pk}A\right)\mathrm{vec}X = \left(B_k^{\mathrm{T}} \otimes A\right)\mathrm{vec}X,$$

所以

$$\mathrm{vec}(AXB) = \begin{bmatrix} B_1^{\mathrm{T}} \otimes A \\ \vdots \\ B_q^{\mathrm{T}} \otimes A \end{bmatrix}\mathrm{vec}X = \left(B^{\mathrm{T}} \otimes A\right)\mathrm{vec}X. \qquad \square$$

通常,对于许多线性矩阵变换和线性矩阵方程,Kronecker 积可给它们以方便的表示.该引理是很关键的.

7.1.11 推论　设 $A, B, C, A_1, \cdots, A_k, B_1, \cdots, B_k, X, Y \in \mathbb{C}^{n\times n}$.

(1)　$AX = B \Leftrightarrow (I \otimes A)\mathrm{vec}X = \mathrm{vec}B$.

(2)　$AX + XB = C \Leftrightarrow \left((I \otimes A) + \left(B^{\mathrm{T}} \otimes I\right)\right)\mathrm{vec}X = \mathrm{vec}C$.

(3)　$A_1XB_1 + \cdots + A_kXB_k = C$

$$\Leftrightarrow \left(B_1^{\mathrm{T}} \otimes A_1 + \cdots + B_k^{\mathrm{T}} \otimes A_k\right)\mathrm{vec}X = \mathrm{vec}C.$$

(4)　$AX + YB = C$

$$\Leftrightarrow (I \otimes A)\mathrm{vec}X + \left(B^{\mathrm{T}} \otimes I\right)\mathrm{vec}Y = \mathrm{vec}C.$$

证明留作练习.

7.1.10 和 **7.1.11** 提供了非常有用的矩阵方程的预处理.

7.1.12 引理 设 $T:\mathbb{C}^{m\times n}\to\mathbb{C}^{p\times q}$ 是线性变换.则存在唯一的矩阵 $K(T)\in\mathbb{C}^{pq\times mn}$,使得

$$\mathrm{vec}T(X)=K(T)\mathrm{vec}X,\ \forall X\in\mathbb{C}^{m\times n}.$$

证 因为 $X\to\mathrm{vec}X$ 的映射 $\mathrm{vec}:\mathbb{C}^{m\times n}\to\mathbb{C}^{mn}$ 是同构映射,$T(X)\to\mathrm{vec}T(X)$ 的映射 $\mathrm{vec}:\mathbb{C}^{p\times q}\to\mathbb{C}^{pq}$ 也是同构映射,所以线性映射 $T:\mathbb{C}^{m\times n}\to\mathbb{C}^{p\times q}$ 等价于映射

$$\tilde{T}:\mathbb{C}^{mn}\to\mathbb{C}^{pq};\ \tilde{T}(\mathrm{vec}X)\equiv\mathrm{vec}T(X),\forall X\in\mathbb{C}^{m\times n},$$

而且 $\tilde{T}$ 也是线性的.于是,依 **1.4.1**,即得定理的结论. □

7.1.13 定义 线性映射 $T:\mathbb{C}^{n\times n}\to\mathbb{C}^{n\times n}$ 称为**求导映射**,如果满足

$$T(XY)=T(X)Y+XT(Y),\ \forall X,Y\in\mathbb{C}^{n\times n}.\qquad(1.11)$$

7.1.14 定理 设 $T:\mathbb{C}^{n\times n}\to\mathbb{C}^{n\times n}$ 是线性映射.T 为求导映射的充分必要条件是存在 $C\in\mathbb{C}^{n\times n}$,使得

$$T(X)=CX-XC,\ \forall X\in\mathbb{C}^{n\times n}.\qquad(1.12)$$

证 充分性.由(1.12)有

$$\begin{aligned}T(XY)&=CXY-XYC\\&=CXY-XCY+XCY-XYC\\&=(CX-XC)Y+X(CY-YC)\\&=T(X)Y+XT(Y),\quad\forall X,Y\in\mathbb{C}^{n\times n}.\end{aligned}$$

必要性.(1.11)等价于

$$\mathrm{vec}T(XY)=\mathrm{vec}(T(X)Y)+\mathrm{vec}(XT(Y)),$$
$$\forall X,Y\in\mathbb{C}^{n\times n}.\qquad(1.13)$$

依 **7.1.12**,存在唯一的 $K(T)\in\mathbb{C}^{n^2\times n^2}$,使得

$$\mathrm{vec}T(X)=K(T)\mathrm{vec}X,\ \forall X\in\mathbb{C}^{n\times n}.$$

再注意到

$$\mathrm{vec}(AB)=(I\otimes A)\mathrm{vec}B,\ \forall A,B\in\mathbb{C}^{n\times n},\qquad(1.14)$$

于是有

$$\mathrm{vec}T(XY)=K(T)\mathrm{vec}(XY)=K(T)(I\otimes X)\mathrm{vec}Y,$$
$$\mathrm{vec}(T(X)Y)=(I\otimes T(X))\mathrm{vec}Y,$$
$$\mathrm{vec}(XT(Y))=(I\otimes X)\mathrm{vec}T(Y)=(I\otimes X)K(T)\mathrm{vec}Y.$$

(1.13)可以改写成

$$K(T)(I\otimes X)\mathrm{vec}Y=(I\otimes T(X))\mathrm{vec}Y+(I\otimes X)K(T)\mathrm{vec}Y,$$
$$\forall X,Y\in\mathbb{C}^{n\times n}.$$

由此推出

$$K(T)(I\otimes X)-(I\otimes X)K(T)=I\otimes T(X),$$
$$\forall X\in\mathbb{C}^{n\times n}.\qquad(1.15)$$

将$K(T)$写成分块形式

$$K(T)=[K_{ij}];\quad K_{ij}\in\mathbb{C}^{n\times n},\ i,j=1,\cdots,n.$$

(1.15)等价于

$$\begin{cases}T(X)=K_{ii}X-XK_{ii}, & i=1,\cdots,n,\\ K_{ij}X-XK_{ij}=0, & i,j=1,\cdots,n,\quad i\neq j,\end{cases}$$
$$\forall X\in\mathbb{C}^{n\times n}.\ (1.16)$$

特别

$$T(X)=K_{11}X-XK_{11},\ \forall X\in\mathbb{C}^{n\times n}.$$

这就是关于T的形如(1.12)的表示. □

7.1.15 定理 对于任意给定的正整数m和n,存在唯一的矩阵$P(m,n)\in\mathbb{C}^{mn\times mn}$,使得

$$\mathrm{vec}X^{\mathrm{T}}=P(m,n)\mathrm{vec}X,\ \forall X\in\mathbb{C}^{m\times n}.\qquad(1.17)$$

如此矩阵$P(m,n)$仅依赖于维数m和n,具体地

$$P(m,n)=\sum_{i=1}^{m}\sum_{j=1}^{n}E_{ij}\otimes E_{ij}^{\mathrm{T}}=\begin{bmatrix}E_{11}^{\mathrm{T}} & E_{12}^{\mathrm{T}} & \cdots & E_{1n}^{\mathrm{T}}\\ E_{21}^{\mathrm{T}} & E_{22}^{\mathrm{T}} & \cdots & E_{2n}^{\mathrm{T}}\\ \cdots & \cdots & \cdots & \cdots\\ E_{n1}^{\mathrm{T}} & E_{n2}^{\mathrm{T}} & \cdots & E_{nn}^{\mathrm{T}}\end{bmatrix}, \quad (1.18)$$

其中每个 $E_{ij}\in\mathbb{C}^{m\times n}$ 在 (i,j) 位置的元素为1,其余元素为0.而且 $P(m,n)$ 是置换矩阵,并有

$$P(m,n)=P(n,m)^{\mathrm{T}}=P(n,m)^{-1}. \quad (1.19)$$

证 对于任意的 $X=[x_{ij}]\in\mathbb{C}^{m\times n}$,有

$$E_{ij}^{\mathrm{T}}XE_{ij}^{\mathrm{T}}=x_{ij}E_{ij}^{\mathrm{T}},\quad i=1,\cdots,m,\ j=1,\cdots,n.$$

于是

$$X^{\mathrm{T}}=\sum_{i=1}^{m}\sum_{j=1}^{n}x_{ij}E_{ij}^{\mathrm{T}}=\sum_{i=1}^{m}\sum_{j=1}^{n}E_{ij}^{\mathrm{T}}XE_{ij}^{\mathrm{T}}.$$

再利用 **7.1.10**,得

$$\mathrm{vec}X^{\mathrm{T}}=\sum_{i=1}^{m}\sum_{j=1}^{n}\mathrm{vec}\left(E_{ij}^{\mathrm{T}}XE_{ij}^{\mathrm{T}}\right)=\sum_{i=1}^{m}\sum_{j=1}^{n}\left(E_{ij}\otimes E_{ij}^{\mathrm{T}}\right)\mathrm{vec}X$$

(1.17)和(1.18)得证.

因为

$$\left(X^{\mathrm{T}}\right)^{\mathrm{T}}=X,\quad X^{\mathrm{T}}\in\mathbb{C}^{n\times m},$$

有

$$\mathrm{vec}X=P(n,m)\mathrm{vec}X^{\mathrm{T}}=P(n,m)P(m,n)\mathrm{vec}X,$$

所以

$$P(m,n)=P(n,m)^{-1}.$$

另一方面,用 $\tilde{E}_{ij}$ 表示 $\mathbb{C}^{n\times m}$ 中 (i,j) 位置的元素为1,其余元素为0的矩阵,注意到

$$\tilde{E}_{ij}=\left(E_{ji}\right)^{\mathrm{T}}=E_{ji}^{\mathrm{T}}$$

有

$$P(n,m)=\sum_{i=1}^{n}\sum_{j=1}^{m}\tilde{E}_{ij}\otimes\tilde{E}_{ij}^{\mathrm{T}}=\sum_{i=1}^{n}\sum_{j=1}^{m}E_{ji}^{\mathrm{T}}\otimes E_{ji}$$
$$=\sum_{i=1}^{m}\sum_{j=1}^{n}\left(E_{ij}\otimes E_{ij}^{\mathrm{T}}\right)^{\mathrm{T}}=P(m,n)^{\mathrm{T}},$$

(1.19)得证.

最后,因 $P(m,n)$ 的元素为 0 或 1,而且 $P(m,n)^{-1}=P(m,n)^{\mathrm{T}}$,故必是置换矩阵. □

7.1.16 推论 设 m,n,p,q 是正整数,$P(p,m)\in\mathbb{C}^{mp\times mp}$ 和 $P(n,q)\in\mathbb{C}^{nq\times nq}$ 表示由(1.18)定义的置换矩阵.则

$$B\otimes A=P(m,p)^{\mathrm{T}}(A\otimes B)P(n,q),\quad \forall A\in\mathbb{C}^{m\times n},B\in\mathbb{C}^{p\times q}. \tag{1.20}$$

这就是说,$B\otimes A$ 总是**置换等价**于 $A\otimes B$.

当 $m=n$ 和 $p=q$ 时,

$$B\otimes A=P(n,p)^{\mathrm{T}}(A\otimes B)P(n,p),\quad \forall A\in\mathbb{C}^{n\times n},B\in\mathbb{C}^{p\times p}.$$

也就是说,当 A 和 B 都是方阵时,$B\otimes A$ 总是**置换相似**于 $A\otimes B$,而且置换相似性仅依赖维数 n 和 p.

更一般地,设 $A_1,\cdots,A_k\in\mathbb{C}^{m\times n}$,$B_1,\cdots,B_k\in\mathbb{C}^{p\times q}$,则

$$B_1\otimes A_1+\cdots+B_k\otimes A_k$$
$$=P(m,p)^{T}(A_1\otimes B_1+\cdots+A_k\otimes B_k)P(n,q). \tag{1.21}$$

证 考虑线性映射

$$T:\mathbb{C}^{n\times q}\to\mathbb{C}^{m\times p};\ T(X)=AXB^{\mathrm{T}},\forall X\in\mathbb{C}^{n\times q},$$

则 $T(X)^{\mathrm{T}}=BX^{\mathrm{T}}A^{\mathrm{T}}$.依 **7.1.10**,$\forall X\in\mathbb{C}^{n\times q}$,有

$$\mathrm{vec}T(X)=\mathrm{vec}\left(AXB^{\mathrm{T}}\right)=(B\otimes A)\mathrm{vec}X \tag{1.22}$$

和

$$\mathrm{vec}T(X)^{\mathrm{T}} = \mathrm{vec}\left(BX^{\mathrm{T}}A^{\mathrm{T}}\right) = (A \otimes B)\mathrm{vec}X^{\mathrm{T}}. \tag{1.23}$$

利用 **7.1.15**,有

$$\mathrm{vec}T(X)^{\mathrm{T}} = P(m,p)\mathrm{vec}T(X), \quad \mathrm{vec}X^{\mathrm{T}} = P(n,q)\mathrm{vec}X.$$

(1.23)可写成

$$P(m,p)\mathrm{vec}T(X) = (A \otimes B)P(n,q)\mathrm{vec}X,$$

等价于

$$\mathrm{vec}T(X) = P(m,p)^{T}(A \otimes B)P(n,q)\mathrm{vec}X.$$

比较此式与(1.22),即得(1.20). □

7.1.17 定义 设 $A \in \mathbb{C}^{n\times n}, B \in \mathbb{C}^{m\times m}$. $mn \times mn$ 矩阵

$$(I_m \otimes A) + (B \otimes I_n), \tag{1.24}$$

称为 A 与 B 的 **Kronecker 和**.

这一概念基于形如(1.24)的矩阵自然地出现在矩阵方程的 Kronecker 积的表示形式中.为了给出具体的说明,考虑矩阵方程

$$AX + XB = C, \tag{1.25}$$

其中

$$A \in \mathbb{C}^{n\times n}, B \in \mathbb{C}^{m\times m}, C, X \in \mathbb{C}^{n\times m}.$$

此方程包括若干重要的特殊情形:诸如 Lyapunov 方程

$$XA + A^{*}X = C, \quad A, C, X \in \mathbb{C}^{n\times n}. \tag{1.26}$$

交换性方程

$$AX = XA, \quad A, X \in \mathbb{C}^{n\times n}. \tag{1.27}$$

方程(1.25)的 Kronecker 积的表示形式为

$$\left((I_m \otimes A) + (B^{T} \otimes I_n)\right)\mathrm{vec}X = \mathrm{vec}C, \tag{1.28}$$

其系数矩阵就是 A 与 B^{T} 的 Kronecker 和.

7.1.18 定理 设 $A \in \mathbb{C}^{n\times n}, B \in \mathbb{C}^{m\times m}$.则

(1) $I_m \otimes A$ 与 $B \otimes I_n$ 可交换.

(2) 如果 $\lambda \in \lambda(A)$ 且 $x \in \mathbb{C}^{n}$ 是相应的 A 的特征向量,而 $\mu \in \lambda(B)$ 且 $y \in \mathbb{C}^{m}$ 是相应的 B 的特征向量,那么 $\lambda + \mu$ 是

Kronecker 和$(I_m \otimes A)+(B \otimes I_n)$的特征值,并且$y \otimes x \in \mathbb{C}^{nm}$是相应的特征向量.

(3) Kronecker 和的每个特征值可以表示为A与B的特征值的和.也就是说,如果

$$\lambda(A)=\{\lambda_1,\cdots,\lambda_n\},\lambda(B)=\{\mu_1,\cdots,\mu_m\},$$

则

$$\lambda((I_m \otimes A)+(B \otimes I_n))=\{\lambda_i+\mu_j : i=1,\cdots,n; j=1,\cdots,m\}. \tag{1.29}$$

这里三者均按代数重数计.

(4) 成立

$$\lambda((I_m \otimes A)+(B \otimes I_n))=\lambda((I_n \otimes B)+(A \otimes I_m)). \tag{1.30}$$

证 利用混合积性质(1.7),有

$$\begin{aligned}(I_m \otimes A)(B \otimes I_n) &= I_m B \otimes A I_n \\ &= B \otimes A=(B \otimes I_n)(I_m \otimes A).\end{aligned}$$

因此$I_m \otimes A$与$B \otimes I_n$可交换.

现在假设

$$Ax=\lambda x,\quad By=\mu y,\quad x \neq 0,\quad y \neq 0,$$

则有

$$\begin{aligned}&((I_m \otimes A)+(B \otimes I_n))(y \otimes x) \\ &\quad=(I_m y \otimes Ax)+(By \otimes I_n x) \\ &\quad=(y \otimes \lambda x)+(\mu y \otimes x)=(\lambda+\mu)(y \otimes x).\end{aligned}$$

表明$\lambda+\mu$与$y \otimes x$分别是 Kronecker 和的特征值与相应的特征向量.

Schur 上三角标准形定理 **5.3.3** 保证存在酉矩阵$U \in \mathbb{C}^{n\times n}$与$V \in \mathbb{C}^{m\times m}$,使得

$$U^* A U=T_A,$$

与

$$V^* B V=T_B,$$

其中 T_A 与 T_B 是上三角矩阵.由此并利用(1.7),有

$$(V\otimes U)^*(I_m\otimes A)(V\otimes U)=(V^*I_mV)\otimes(U^*AU)$$

$$=I_m\otimes T_A=\begin{bmatrix} T_A & & & 0 \\ & T_A & & \\ 0 & & \ddots & \\ & & & T_A \end{bmatrix}\in\mathbb{C}^{nm\times nm},$$

和

$$(V\otimes U)^*(B\otimes I_n)(V\otimes U)=(V^*BV)\otimes(U^*I_nU)$$

$$=T_B\otimes I_n=\begin{bmatrix} \mu_1 I_n & * & \cdots & * \\ & \mu_2 I_n & \ddots & \vdots \\ & 0 & \ddots & * \\ & & & \mu_m I_n \end{bmatrix}\in\mathbb{C}^{nm\times nm},$$

均为上三角矩阵.因此

$$(V\otimes U)^*((I_m\otimes A)+(B\otimes I_n))(V\otimes U)$$
$$=(I_m\otimes T_A)+(T_B\otimes I_n)$$

是上三角矩阵.

注意到 $V\otimes U\in\mathbb{C}^{nm\times nm}$ 是酉矩阵, $(I_m\otimes T_A)+(T_B\otimes I_n)$ 的对角元素必是 A 与 B 的 Kronecker 和的特征值,且表明它们是 T_B 的每个对角元素与 T_A 的所有对角元素配对之和. □

7.1.19 定理 设 $A\in\mathbb{C}^{n\times n}$, $B\in\mathbb{C}^{m\times m}$. 方程(1.25)即

$$AX+XB=C$$

对每个 $C\in\mathbb{C}^{n\times m}$ 有唯一解 $X\in\mathbb{C}^{n\times m}$ 的充分必要条件是

$$\lambda(A)\cap\lambda(-B)=\phi. \tag{1.31}$$

证 注意到 B^{T} 与 B 有相同的特征值,而且方程(1.25)的 Kronecker 积的表示形式是方程(1.28).依 **7.1.18**,即知(1.28)的系数矩阵有零特征值的充分必要条件是

$$\lambda(A)\cap\lambda(-B)\neq\emptyset.$$

因此,方程(1.25)对 X 而言,非奇异(即有唯一解 $X\in\mathbb{C}^{n\times m}$)的充分必要条件是(1.31). □

7.1.20 定理 设 $A\in\mathbb{C}^{n\times n}$.方程(1.26)即

$$XA+A^{*}X=C$$

对每个 $C\in\mathbb{C}^{n\times n}$ 有唯一解 $X\in\mathbb{C}^{n\times n}$ 的充分必要条件是

$$\lambda(A)\cap\overline{\lambda(-A)}=\emptyset, \tag{1.32}$$

也就是说,$\lambda\in\lambda(A)$ 蕴涵 $-\overline{\lambda}\notin\lambda(A)$.

特别,若 A 是正稳定的,则方程(1.26)必有唯一解 X;而且当 C 为自伴矩阵时,X 也是自伴矩阵.

证 依 **7.1.19**,方程(1.26)有唯一解的充分必要条件是

$$\lambda(A^{*})\cap\lambda(-A)=\emptyset.$$

由此,以及 $\lambda(A^{*})=\overline{\lambda(A)}$,而且 $\lambda(-A)=-\lambda(A)$,即得(1.32).

若 A 是正稳定的,即其所有特征值有正实部,则显然成立(1.32),故方程(1.26)总有唯一解.另外,注意到

$$X^{*}A+A^{*}X^{*}=C^{*}, \tag{1.33}$$

因此,当 A 是正稳定时,如果 C 是自伴的,即 $C^{*}=C$,那么由解的唯一性知,(1.33)的解 X^{*} 就是(1.26)的解 X,即 $X^{*}=X$,故 X 也是自伴的. □

关于 Kronecker 积的进一步的讨论可参看[HJ1985].

7.2 Hadamard 积

7.2.1 定义 设 $A=[a_{ij}],B=[b_{ij}]\in\mathbb{C}^{m\times n}$.用 $A\circ B$ 表示 A 和 B 的对应元素相乘而成的 $m\times n$ 矩阵:

$$A \circ B \equiv \begin{bmatrix} a_{11}b_{11} & a_{12}b_{12} & \cdots & a_{1n}b_{1n} \\ a_{21}b_{21} & a_{22}b_{22} & \cdots & a_{2n}b_{2n} \\ \cdots & \cdots & \cdots & \cdots \\ a_{m1}b_{m1} & a_{m2}b_{m2} & \cdots & a_{mn}b_{mn} \end{bmatrix} \tag{2.1}$$

$A \circ B$ 称为 A 和 B 的 **Hadamard 积**,也称为 **Schur 积**.

不巧,这种乘法的记号 $\circ$ 和映射复合的记号相重,因此,在容易引起误会的场合使用时应加以申明.

如此 Hadamard 乘法远比通常矩阵乘法简单,但未被广泛地了解.它出现在广泛而多样的方方面面之中,诸如周期函数卷积的三角矩阵,积分方程核的积,偏微分方程中的弱极小原理,概率论中的特征函数,组合论中的结合方案研究,算子理论中关于无限矩阵的 Hadamard 积等.

Hadamard 积的可相乘条件是只要两个矩阵有相同的行数和相同的列数.

显然,如此乘法与通常矩阵乘法不同,它是可交换的,即

$$A \circ B = B \circ A. \tag{2.2}$$

7.2.2 定理 设 $A, B, C \in \mathbb{C}^{m\times n}$.则成立

(1) $A \circ (B + C) = A \circ B + A \circ C$.

(2) $A \circ (B \circ C) = (A \circ B) \circ C$.

(3) $(A \circ B)^{\mathrm{T}} = A^{\mathrm{T}} \circ B^{\mathrm{T}}$.

(4) $(A \circ B)^* = A^* \circ B^*$.

(5) 如果 A 和 B 是自伴矩阵,那么 $A \circ B$ 也是自伴矩阵.

(6) 如果 A 和 B 是斜自伴矩阵,那么 $A \circ B$ 是自伴矩阵.

(7) 如果 A 是自伴矩阵,B 是斜自伴矩阵,那么 $A \circ B$ 是斜自伴矩阵.

这些基本性质可直接由 Hadamard 积的定义推出,留作练习.

7.2.3 引理 设 $A, B \in \mathbb{C}^{m\times n}$.则成立

$$A \circ B = (A \otimes B)(\alpha, \beta), \tag{2.3}$$

其中

$$\alpha = \{1, \quad m+2, \quad 2m+3, \quad \cdots, \quad m^2\}$$

和

$$\beta = \{1, \quad n+2, \quad 2n+3, \quad \cdots, \quad n^2\}$$

是指标集,$(A \otimes B)(\alpha, \beta)$是 $A \otimes B$ 的由 α 指定的行和 β 指定的列所确定的子矩阵.

特别,当 $m = n$ 时,$A \circ B$ 是 $A \otimes B$ 的主子矩阵.

证 检验等式(2.3)右端 $A \otimes B$ 的子矩阵的元素即可. □

这一引理表明,Hadamard 积 $A \circ B$ 等同于 Kronecker 积 $A \otimes B$ 的一个子矩阵.

7.2.4 引理 设 $A, B \in \mathbb{C}^{m\times n}$,且设 $D \in \mathbb{C}^{m\times m}$ 和 $E \in \mathbb{C}^{n\times n}$ 是对角矩阵.则

$$\begin{aligned} D(A \circ B)E &= (DAE) \circ B = (DA) \circ (BE) \\ &= (AE) \circ (DB) = A \circ (DBE). \end{aligned} \tag{2.4}$$

这一引理表明,Hadamard 乘法同对角矩阵作通常乘法是可交换的.证明留作练习.

7.2.5 引理 设 $A, B \in \mathbb{C}^{m\times n}$,记号

$$D_x \equiv \operatorname{diag}(x_1, \cdots, x_n), \quad \forall x = (x_1, \cdots, x_n)^{\mathrm{T}} \in \mathbb{C}^n, \tag{2.5}$$

则

$$\left(AD_xB^{\mathrm{T}}\right)_{ii} = \left((A \circ B)x\right)_i, \quad i = 1, \cdots, m. \tag{2.6}$$

这里,等式左边是矩阵的第 i 个对角元素,等式右边是向量的第 i 个分量.

证 如果 $A=[a_{ij}],B=[b_{ij}]$,那么

$$\left(AD_xB^{\mathrm{T}}\right)_{ii}=\sum_{j=1}^{n}a_{ij}x_jb_{ij}=\sum_{j=1}^{n}a_{ij}b_{ij}x_j=\left(\left(A\circ B\right)x\right)_i. \qquad \square$$

7.2.6 引理 设 $A,B,C\in\mathbb{C}^{m\times n}$.则

$$\left(\left(A\circ B\right)C^{\mathrm{T}}\right)_{ii}=\left(\left(A\circ C\right)B^{\mathrm{T}}\right)_{ii},\quad i=1,\cdots,m \tag{2.7}$$

这是三重混合积的一个性质.证明留作练习.

7.2.7 引理 设 $A,B\in\mathbb{C}^{m\times n},y\in\mathbb{C}^{m},x\in\mathbb{C}^{n}$.则

$$y^{*}\left(A\circ B\right)x=\operatorname{tr}\left(D_y^{*}AD_xB^{\mathrm{T}}\right) \tag{2.8}$$

其中记号 D_x 和 D_y 由(2.5)定义.

证 令 $e=\left(1,\cdots,1\right)^{\mathrm{T}}\in\mathbb{C}^{m}$.

注意到 $D_ye=y$,有

$$\begin{aligned}y^{*}\left(A\circ B\right)x&=e^{\mathrm{T}}D_y^{*}\left(A\circ B\right)x\\&=e^{\mathrm{T}}\left(\left(D_y^{*}A\circ B\right)x\right)=\operatorname{tr}\left(D_y^{*}AD_xB^{\mathrm{T}}\right)\end{aligned}$$

第二个等号依据 **7.2.4**;第三个等号利用了 **7.2.5**,用 $D_y^{*}A$ 替代那里的 A. $\square$

7.2.8 定理 设 $A,B\in\mathbb{C}^{m\times n}$.则

$$\operatorname{rank}\left(A\circ B\right)\le\left(\operatorname{rank}A\right)\left(\operatorname{rank}B\right) \tag{2.9}$$

证 任何秩为 r 的矩阵可以写成 r 个秩 1 矩阵之和,其每个秩 1 矩阵是两个向量的外积(参见[HJ1985]).因此,若

$$\operatorname{rank}A=r_1,\quad \operatorname{rank}B=r_2,$$

则有

$$A=\sum_{i=1}^{r_1} x_i y_i^*,\quad B=\sum_{j=1}^{r_2} u_j v_j^*,$$

其中$x_i,u_j\in\mathbb{C}^m, y_i,v_j\in\mathbb{C}^n, i=1,\cdots,r_1, j=1,\cdots,r_2$.于是

$$A\circ B=\sum_{i=1}^{r_1}\sum_{j=1}^{r_2}\left(x_i\circ u_j\right)\left(y_i\circ v_j\right)^*.$$

这表明$A\circ B$至多是r_1r_2个秩1矩阵之和.因此

$$\operatorname{rank}\left(A\circ B\right)\le r_1r_2=\left(\operatorname{rank}A\right)\left(\operatorname{rank}B\right).\qquad\square$$

7.2.9 Schur 积定理 (1) 若$A,B\in\mathbb{C}^{n\times n}$是半正定矩阵,则$A\circ B$也是半正定矩阵.

(2) 若B是正定矩阵,A是半正定矩阵且没有零对角元素,则$A\circ B$是正定矩阵.

(3) 若A和B都是正定矩阵,则$A\circ B$也是正定矩阵.

证 先证(2).

利用**7.2.7**,有

$$x^*\left(A\circ B\right)x=\operatorname{tr}\left(D_x^*AD_xB^{\mathrm{T}}\right),\ \forall x\in\mathbb{C}^n.\qquad(2.10)$$

注意,一个自伴矩阵和一个正定矩阵乘积的惯性相同于自伴矩阵的惯性(参见[HJ1985]的定理 7.6.3).对于任何$x\in\mathbb{C}^n$,矩阵$D_x^*AD_x$是半正定的;如果$x\neq 0$且A没有对角元素为零,那么$D_x^*AD_x\neq 0$,因而它至少有一个正特征值.又若B是正定的,则B^{T}也是正定的.

因此,$\left(D_x^*AD_x\right)B^{\mathrm{T}}$所有特征值是非负的,而且至少有一个是正的.由此推出

$$\operatorname{tr}\left(D_x^*AD_xB^{\mathrm{T}}\right)>0,$$

于是,从(2.10)即知$A\circ B$是正定的.

现在将结论(2)应用于

$$A_\varepsilon\equiv A+\varepsilon I,\ B_\varepsilon\equiv B+\varepsilon I,\ \varepsilon>0,$$

并令$\varepsilon\to 0$,即可推出(1).

最后,基于结论(2)及如下事实:若 $A=[a_{ij}]$ 正定,则

$$a_{ii}=e_i^* Ae_i>0,\ i=1,\cdots,n.$$

即得结论(3). □

这一定理表明正定矩阵类和半正定矩阵类在 Hadamard 积运算下是封闭的.

这是一个定性的结果.还有各种定量的结果.

7.2.10 定理 设 $A,B\in\mathbb{C}^{n\times n}$ 是半正定矩阵.则成立

$$\lambda_{\min}(A\circ B)\geq\lambda_{\min}(A)\lambda_{\min}(B) \tag{2.11}$$

和

$$\lambda_{\max}(A\circ B)\leq\lambda_{\max}(A)\lambda_{\max}(B) \tag{2.12}$$

其中 $\lambda_{\min}(A)$ 和 $\lambda_{\max}(A)$ 分别表示 A 的最小特征值和最大特征值.

证 依 **7.2.3**, $A\circ B$ 是 Kronecker 积 $A\otimes B$ 的主子矩阵.

因为 $A\otimes B$ 的特征值正好是 A 和 B 的特征值的两两之积(见 **7.1.7**),其最小特征值是

$$\lambda_{\min}(A)\lambda_{\min}(B).$$

最大特征值是

$$\lambda_{\max}(A)\lambda_{\max}(B).$$

而自伴矩阵的任何主子矩阵的特征值又必处在其最小特征值和最大特征值之间(见[HJ1985]的定理 4.3.15),因此成立(2.11)和(2.12). □

此定理是一个比较弱的定量估计.例如,A 正定,$B=A^{-1}$ 的情形,按照(2.11),有

$$\lambda_{\min}(A\circ A^{-1})\geq\lambda_{\min}(A)\lambda_{\min}(A^{-1})=\lambda_{\min}(A)/\lambda_{\max}(A),$$

得到的是一个很差的下界;而依下述定理就可得到好得多的可用的下界 1.

7.2.11 定理 设 $A,B\in\mathbb{C}^{n\times n}$ 是半正定矩阵.则成立

$$\lambda_{\min}(A\circ B)\geq\lambda_{\min}\left(AB^{\mathrm{T}}\right) \tag{2.13}$$

和

$$\lambda_{\min}(A\circ B)\geq\lambda_{\min}(AB). \tag{2.14}$$

证明从略.可参见[HJ1991]的定理 5.3.1.

7.2.12 定理 设 $A,B\in\mathbb{C}^{n\times n}$ 是自伴矩阵,而且 $A=[a_{ij}]$ 是半正定的.则 $A\circ B$ 的任何特征值 $\lambda_{A\circ B}$ 有估计式

$$\begin{aligned}\lambda_{\min}(A)\lambda_{\min}(B)&\leq\lambda_{\min}(B)\min_{1\leq i\leq n}a_{ii}\\&\leq\lambda_{A\circ B}\\&\leq\lambda_{\max}(B)\max_{1\leq i\leq n}a_{ii}\leq\lambda_{\max}(A)\lambda_{\max}(B).\end{aligned} \tag{2.15}$$

证 因 $B-\lambda_{\min}(B)I$ 和 A 是半正定的,故 $A\circ(B-\lambda_{\min}(B)I)$ 也是半正定的.

设 $x=(x_1,\cdots,x_n)^{\mathrm{T}}$ 是 $A\circ B$ 的相应特征值 $\lambda_{A\circ B}$ 的单位特征向量,则有

$$\begin{aligned}0&\leq x^*\left[A\circ(B-\lambda_{\min}(B)I)\right]x\\&=x^*(A\circ B)x-\lambda_{\min}(B)x^*(A\circ I)x\\&=\lambda_{A\circ B}-\lambda_{\min}(B)\sum_{i=1}^{n}|x_i|^2a_{ii}\leq\lambda_{A\circ B}-\lambda_{\min}(B)\min_{1\leq i\leq n}a_{ii}.\end{aligned}$$

这就完成了下界的证明.

上界的推导是类似的,留作练习. □

7.2.13 引理 设 $B=[b_{ij}]\in\mathbb{C}^{n\times n}$ 是可对角化矩阵,其特征值为 $\lambda_1,\cdots,\lambda_n$,而且非奇异矩阵 $A\in\mathbb{C}^{n\times n}$ 使得

$$B=A\operatorname{diag}(\lambda_1,\cdots,\lambda_n)A^{-1} \tag{2.16}$$

则 B 的对角元素和特征值之间成立如下关系:

$$\begin{bmatrix} b_{11} \\ b_{22} \\ \vdots \\ b_{nn} \end{bmatrix} = \left(A \circ (A^{-1})^{\mathrm{T}}\right) \begin{bmatrix} \lambda_1 \\ \lambda_2 \\ \vdots \\ \lambda_n \end{bmatrix}. \tag{2.17}$$

证 从(2.16)直接计算 B 的对角元素. □

当 A 是酉矩阵时,(2.16)确定的 B 是正规矩阵,而且

$$A \circ \left(A^{-1}\right)^{\mathrm{T}} = A \circ \overline{A}$$

是双随机矩阵(见 **4.5.1**).如果还有 B 的特征值 $\lambda_1, \cdots, \lambda_n$ 是实的,那么 B 还是自伴的,从(2.17)推出 B 的对角元素优化于其特征值(见 **4.5.8**).

在化学工程性能设计的一种途径中,广泛使用形如 $A \circ \left(A^{-1}\right)^{\mathrm{T}}$ 矩阵.在那里,**增益矩阵**(gain matrix) $A \in \mathbb{R}^{n \times n}$ 是非奇异的,它描述化学过程中输入和输出之间的关系,并把矩阵 $A \circ \left(A^{-1}\right)^{\mathrm{T}}$ 称为**相对增益阵列**(relative gain array).设计者要检验在设备设计中确定输入和输出目标的相对增益矩阵.

7.2.14 引理 对任一非奇异矩阵 $A \in \mathbb{C}^{n \times n}$,定义

$$\Phi(A) \equiv A \circ A^{-1}, \quad \Phi_{\mathrm{T}}(A) \equiv A \circ \left(A^{-1}\right)^{\mathrm{T}}, \tag{2.18}$$

则成立

(1) $\Phi_{\mathrm{T}}(A)$ 的各行的和与各列的和均为1.

(2) 对任何非奇异对角矩阵 $D, E \in \mathbb{C}^{n \times n}$,有

$$\Phi_{\mathrm{T}}(DAE) = \Phi_{\mathrm{T}}(A)$$

和

$$\Phi(DAE) = \left(D^{-1}E\right)^{-1} \Phi(A) \left(D^{-1}E\right).$$

(3) 对任何置换矩阵 $P, Q \in \mathbb{C}^{n \times n}$,有

$$\Phi_{\mathrm{T}}(PAQ) = P\Phi_{\mathrm{T}}(A)Q.$$

(4) 对任何置换矩阵 $P \in \mathbb{C}^{n\times n}$,有

$$\Phi_{\mathrm{T}}(P) = P.$$

(5) $\Phi(A^{-1}) = \Phi(A)$.

(6) $\Phi(A^{\mathrm{T}}) = (\Phi(A))^{\mathrm{T}}$.

(7) $\Phi_{\mathrm{T}}(A^{-1}) = \Phi_{\mathrm{T}}(A^{\mathrm{T}})$.

证明留作练习.

7.2.15 定理 设 $A \in \mathbb{C}^{n\times n}$ 是正定矩阵.则

$$\lambda_{\min}(\Phi(A)) \geq 1, \quad \lambda_{\min}(\Phi_{\mathrm{T}}(A)) = 1. \tag{2.19}$$

证 (2.19)中:前者从(2.14)推出;后者得自(2.13)以及因 **7.2.14** 的(1)而 $1 \in \lambda(\Phi_{\mathrm{T}}(A))$. □

7.2.16 定理 对于任何 $A, B \in \mathbb{C}^{m\times n}$,成立

$$\sigma_1(A \circ B) \leq \sigma_1(A)\sigma_1(B),$$

其中 $\sigma_1(A)$ 是 A 的最大奇异值.

证明留作练习.

7.3 Fan 积及有关非负矩阵的 Hadamard 积

7.3.1 定义 设 $A = [a_{ij}], B = [b_{ij}] \in \mathbb{C}^{m\times n}$.并设

$$c_{ij} = \begin{cases} a_{ii}b_{ii}, & j = i, \\ -a_{ij}b_{ij}, & j \neq i, \end{cases} \quad i = 1, \cdots, m, j = 1, \cdots, n. \tag{3.1}$$

记 $A \star B \equiv [c_{ij}] \in \mathbb{C}^{m\times n}$,并称其为 A 和 B 的 **Fan 积**(以 Fan Ky

命名).

Fan 积是 Hadamard 积的一种变异.

7.3.2 引理 (1) 若 $A,B\in\mathbb{R}^{n\times n}$ 是 M-矩阵,则 $A\star B$ 也是 M-矩阵.

(2) 若 $A,B\in\mathbb{C}^{n\times n}$ 是H-矩阵,则 $A\star B$ 也是H-矩阵, $A\circ B$ 是非奇异的.

因此,M-矩阵类和 H-矩阵类在 Fan 积$\star$下都是封闭的.

证 (1) $A,B\in Z_n$,显然 $A\star B\in Z_n$.因此只要验证 **4.6.5** 中任一等价条件对 $A\star B$ 成立,便可断定是 M-矩阵.选取 **4.6.5** 的(15).设 D 和 E 是正对角矩阵,使得 AD 和 BE 是行严格对角优势的.注意到对于非负实数 $a,\alpha_1,\cdots,\alpha_m$ 和 $b,\beta_1,\cdots,\beta_m$ 来说,如果

$$a\geq\sum_{i=1}^{m}\alpha_i,\quad b\geq\sum_{i=1}^{m}\beta_i\,,$$

那么

$$ab\geq\left(\sum_{i=1}^{m}\alpha_i\right)\left(\sum_{i=1}^{m}\beta_i\right)\geq\sum_{i=1}^{m}\alpha_i\beta_i.$$

因此,$(AD)\star(BE)$是严格对角优势的.

另一方面,

$$(AD)\star(BE)=(A\star B)(DE). \tag{3.2}$$

这表明对于 $A\star B$ 存在正对角矩阵 DE ,使得 $(A\star B)(DE)$ 是严格对角优势的.

再一次依据 **4.6.5** 的(15), $A\star B$ 是 M-矩阵.

(2) 依 **4.7.2**, A 和 B 的比较矩阵 $\mathscr{M}(A)$和 $\mathscr{M}(B)$是 M-矩阵.根据(1),

$$\mathscr{M}(A\star B)=\mathscr{M}(A)\star\mathscr{M}(B) \tag{3.3}$$

也是M-矩阵.因此$A\star B$是H-矩阵.

现在考虑$A\circ B$.依 **4.6.5** 的(15),存在正对角矩阵F,使得$\mathscr{M}(A\star B)F$是行严格对角优势的.显然,这个F同样使得$(A\circ B)F$是行严格对角优势的.于是,依 **2.2.3**, $(A\circ B)F$,并因此$A\circ B$都是非奇异的. □

7.3.3 引理 设$A,B\in\mathbb{R}^{n\times n}$, $A\geq 0, B\geq 0$,则

(1) $A\circ B\geq 0$,也就是说,非负矩阵类在 Hadamard 积下是封闭的.

(2) $\rho(A\circ B)\leq\rho(A)\rho(B)$.

证 (1)显然.

现在证明(2).利用 Kronecker 积$\otimes$.依 **7.1.7**,

$$\rho(A\otimes B)=\rho(A)\rho(B),$$

因为$A\circ B\geq 0$,且依 **7.2.3**, $A\circ B$是$A\otimes B$的主子矩阵,因此从 **4.1.4** 可以推出

$$\rho(A\circ B)\leq\rho(A\otimes B)=\rho(A)\rho(B). \quad □$$

7.3.4 推论 设$A,B\in\mathbb{R}^{n\times n}$是M-矩阵,则

$$A^{-1}\circ B^{-1}\geq(A\star B)^{-1}, \tag{3.4}$$

并因此

$$\ell(A\star B)\geq\ell(A)\ell(B), \tag{3.5}$$

这里$\ell(A)$是Z-矩阵A的最小特征值(见 **4.6.25**).

证 令

$$R\equiv(I\circ A)-A,\ S\equiv I-(I\circ A)^{-1}A,$$

故有

$$A=(I\circ A)-R=(I\circ A)(I-S).$$

注意到

$$S\equiv[s_{ij}]\geq 0,\quad s_{11}=s_{22}=\cdots=s_{nn}=0,$$

从而

$$\left(I \circ A\right)^{-1} A = I - S \in Z_n .$$

依 **4.6.5** 的(14),对于 A,存在 $x \in \mathbb{R}^n$, $x > 0$,使得 $Ax > 0$;容易看出这个 x 同样使

$$\left(\left(I \circ A\right)^{-1} A\right) x > 0,$$

于是 $\left(I \circ A\right)^{-1} A = I - S$ 是 M-矩阵,并因此依 **4.6.2**,

$$\rho(S) < 1 .$$

同样,对于 B,有表示式

$$B = (I \circ B)(I - T),$$

其中

$$T \equiv [t_{ij}] \geq 0, \ t_{11} = t_{22} = \cdots = t_{nn} = 0 .$$

而且

$$\rho(T) < 1,$$

这样,

$$A \star B = (I \circ A)(I \circ B)(I - S \circ T).$$

并有

$$\rho(S \circ T) \leq \rho(S)\rho(T) < 1,$$

因此

$$\begin{aligned}
(A \star B)^{-1} &= (I - S \circ T)^{-1} (I \circ B)^{-1} (I \circ A)^{-1} \\
&= \left[\sum_{k=0}^{\infty} (S \circ T)^k\right] (I \circ B)^{-1} (I \circ A)^{-1} \\
&\leq \left[\left(\sum_{k=0}^{\infty} S^k\right) \circ \left(\sum_{k=0}^{\infty} T^k\right)\right] (I \circ B)^{-1} (I \circ A)^{-1} \\
&= \left[(I - S)^{-1} \circ (I - T)^{-1}\right] (I \circ B)^{-1} (I \circ A)^{-1} \\
&= \left[(I \circ A)(I - S)\right]^{-1} \circ \left[(I \circ B)(I - T)\right]^{-1} \\
&= A^{-1} \circ B^{-1} .
\end{aligned}$$

最后,由 **4.1.4** 及 **7.3.3**,

$$\rho((A \star B)^{-1}) \leq \rho\left(A^{-1} \circ B^{-1}\right) \leq \rho\left(A^{-1}\right) \rho\left(B^{-1}\right),$$

所以

$$\ell(A\star B)=\rho((A\star B)^{-1})^{-1}$$
$$\geq\rho(A^{-1})^{-1}\rho(B^{-1})^{-1}=\ell(A)\ell(B). \qquad \square$$

7.3.5 定理 设 $A,B\in\mathbb{R}^{n\times n}$ 是 M-矩阵,则 $A\circ B^{-1}$ 也是 M-矩阵.

证 一个矩阵是或不是 M-矩阵不会因左乘、右乘、或同时左乘与右乘以正对角矩阵而改变.由于 $A\in Z_n$ 并具有正对角元素,而且 $B^{-1}\geq 0$,因此 $A\circ B^{-1}\in Z_n$ 并具有正对角元素.

将 **4.6.5** 的(15)应用于 A 和 $B^{\mathrm T}$,存在正对角矩阵 D 和 E,使得 $\hat A\equiv AD$ 行严格对角优势,$\hat B\equiv EB$ 列严格对角优势(即 $B^{\mathrm T}E$ 行严格对角优势).因而,依 **2.2.6**, $\hat B^{-1}$ 必是行元素严格对角优势的.从直接计算看,行严格对角优势矩阵和行元素严格对角优势矩阵的 Hadamard 积 $\hat A\circ\hat B^{-1}$ 必是行严格对角优势的.于是,$\hat A\circ\hat B^{-1}$ 从而 $A\circ B^{-1}$ 都是 M-矩阵. $\square$

7.3.6 引理 设 $A=[a_{ij}]\in\mathbb{R}^{n\times n}$, $A\geq 0$, $\alpha\geq 1$,则

$$\rho\left(A^{(\alpha)}\right)\leq\rho(A)^{\alpha}, \tag{3.6}$$

这里 $A^{(\alpha)}\equiv\left[a_{ij}^{\alpha}\right]$ 是 Hadamard 幂(见 **9.4.1**).

证 对 A 是不可约的情形加以证明就足够了.此时,依 **4.2.4**,存在正特征向量 $x=(x_1,\cdots,x_n)^{\mathrm T}$,使得 $Ax=\rho(A)x$.取正对角矩阵 $D=\mathrm{diag}(x_1,\cdots,x_n)$,则矩阵 $D^{-1}AD$ 的每行之和都等于 $\rho(A)$.因为

$$\rho\left(\left(D^{-1}AD\right)^{(\alpha)}\right)=\rho\left(\left(D^{(\alpha)}\right)^{-1}A^{(\alpha)}D^{(\alpha)}\right)=\rho\left(A^{(\alpha)}\right),$$

而且所考虑的不等式(3.6)对 A 是1次齐次的:即若(3.6)成立,则满足条件

$$k^{\alpha}\rho\left(A^{(\alpha)}\right)=\rho\left((kA)^{(\alpha)}\right)\leq\rho(kA)^{\alpha}=k^{\alpha}\rho(A)^{\alpha},$$
$$\forall k\in\mathbb{R},k\geq 0,$$

所以,不失一般性,可以假定 A 是行随机矩阵.于是

$$A^{(\alpha)} \le A, \quad \forall \alpha \ge 1,$$

并且依 **4.5.2**, $\rho(A)=1$,从而

$$\rho\left(A^{(\alpha)}\right) \le \rho(A) = 1 = \rho(A)^{\alpha}.$$

□

7.3.7 引理 设 $A=[a_{ij}], B=[b_{ij}] \in \mathbb{R}^{n\times n}, A\ge 0, B\ge 0$, 而且 $\alpha\in\mathbb{R}, 0\le\alpha\le 1$,则

$$\rho\left(A^{(\alpha)}\circ B^{(1-\alpha)}\right) \le \rho(A)^{\alpha}\rho(B)^{1-\alpha}. \tag{3.7}$$

证 当 $\rho(A)=0$ 或 $\rho(B)=0$ 或 $\alpha=0$ 或 $\alpha=1$ 时,(3.7)显然成立.再者,只须借助标准的连续性论证,可以假定 A 和 B 是不可约的,而且如同在 **7.3.6** 的证明,还可假定它们都是行随机矩阵.于是, $\rho(A)=\rho(B)=1$.利用 Hölder 不等式(见 **1.8.9**),

$$\sum_{j=1}^{n} a_{ij}^{\alpha} b_{ij}^{1-\alpha} \le \left(\sum_{j=1}^{n} a_{ij}\right)^{\alpha}\left(\sum_{j=1}^{n} b_{ij}\right)^{1-\alpha} = 1^{\alpha}1^{1-\alpha} = 1, \quad i=1,\cdots,n.$$

由此,

$$\begin{aligned}\rho\left(A^{(\alpha)}\circ B^{(1-\alpha)}\right) &\le \left\|A^{(\alpha)}\circ B^{(1-\alpha)}\right\|_{\infty} \\ &\le 1 = \rho(A)^{\alpha}\rho(B)^{1-\alpha}.\end{aligned}$$

□

7.3.8 定理 设 $A_i \in\mathbb{R}^{n\times n}, A_i\ge 0, i=1,\cdots,k; \alpha_1,\cdots,\alpha_k\ge 0$,而且 $\alpha_1+\cdots+\alpha_k\ge 1$,则

$$\rho\left(A_1^{(\alpha_1)}\circ\cdots\circ A_k^{(\alpha_k)}\right) \le \rho(A_1)^{\alpha_1}\cdots\rho(A_k)^{\alpha_k}. \tag{3.8}$$

当 $A_1,\cdots,A_k$ 不可约而且 $\alpha_1+\cdots+\alpha_k=1$ 时,(3.8)成立等号的充分必要条件是存在正对角矩阵 $D_i\in\mathbb{R}^{n\times n}$ 和正数 $\gamma_i, i=2,\cdots,k$, 使得

$$\gamma_i A_i = D_i^{-1}A_1 D_i, \quad i=2,\cdots,k. \tag{3.9}$$

证 利用数学归纳法.因为有 **7.3.6**, $k=1$ 时(3.8)已经成立;而且,

如果$\alpha \equiv \alpha_1 + \cdots + \alpha_k > 1$,那么

$$\rho\left(A_1^{(\alpha_1)} \circ \cdots \circ A_k^{(\alpha_k)}\right) \le \rho\left(\left(A_1^{\left(\frac{\alpha_1}{\alpha}\right)} \circ \cdots \circ A_k^{\left(\frac{\alpha_k}{\alpha}\right)}\right)^{\alpha}\right)$$

$$\le \rho\left(A_1^{\left(\frac{\alpha_1}{\alpha}\right)} \circ \cdots \circ A_k^{\left(\frac{\alpha_k}{\alpha}\right)}\right)^{\alpha},$$

其中

$$\left(\frac{\alpha_1}{\alpha}\right) + \cdots + \left(\frac{\alpha_k}{\alpha}\right) = 1,$$

所以可以只考虑$\alpha_1 + \cdots + \alpha_k = 1$的情形.

现在,可以假定$\alpha_k < 1$;并且按照归纳法,假定不等式(3.8)对于$k-1$个$A_i \ge 0$时成立.定义$B \ge 0$为

$$B^{(1-\alpha_k)} \equiv A_1^{(\alpha_1)} \circ \cdots \circ A_{k-1}^{(\alpha_{k-1})},$$

并令

$$\beta_i \equiv \alpha_j / (1-\alpha_k), j = 1, \cdots, k-1,$$

于是,$\beta_1 + \cdots + \beta_{k-1} = 1$,而且

$$B \equiv A_1^{(\beta_1)} \circ \cdots \circ A_{k-1}^{(\beta_{k-1})}.$$

这样,

$$\begin{aligned}\rho\left(A_1^{(\alpha_1)} \circ \cdots \circ A_k^{(\alpha_k)}\right) &= \rho\left(B^{(1-\alpha_k)} \circ A_k^{(\alpha_k)}\right) \\ &\le \rho(B)^{1-\alpha_k} \rho(A_k)^{\alpha_k} \\ &\le \left[\rho(A_1)^{\beta_1} \cdots \rho(A_{k-1})^{\beta_{k-1}}\right]^{1-\alpha_k} \rho(A_k)^{\alpha_k} \\ &= \rho(A_1)^{\alpha_1} \cdots \rho(A_k)^{\alpha_k},\end{aligned}$$

其中第一个不等号依据 7.3.7,第二个不等号基于归纳法假设.

等号成立的充分必要条件的证明从略. □

7.3.9 推论 设$A \in \mathbb{R}^{n\times n}$, $A \ge 0$,则

$$\rho\left(A^{\left(\frac{1}{2}\right)}\circ\left(A^{\mathrm{T}}\right)^{\left(\frac{1}{2}\right)}\right)\le\rho(A). \tag{3.10}$$

证 依 **7.3.8**,

$$\rho\left(A^{\left(\frac{1}{2}\right)}\circ\left(A^{\mathrm{T}}\right)^{\left(\frac{1}{2}\right)}\right)\le\rho(A)^{\frac{1}{2}}\rho\left(A^{\mathrm{T}}\right)^{\frac{1}{2}},$$

其右边因 $\rho\left(A^{\mathrm{T}}\right)=\rho(A)$ 而可替以 $\rho(A)$. □

7.3.10 定理 设 $A_1,\cdots,A_k\in\mathbb{C}^{n\times n}$ 是 H-矩阵;$\alpha_1,\cdots,\alpha_k\ge 0$,而且 $\alpha_1+\cdots+\alpha_k\ge 1$,则

$$\ell\left(A_1^{(\alpha_1)}\circ\cdots\circ A_k^{(\alpha_k)}\right)\ge\ell(A_1)^{\alpha_1}\cdots\ell(A_k)^{\alpha_k}, \tag{3.11}$$

这里记号 $\ell(A)$ 的定义见 **4.7.6**. 当 $A_1,\cdots,A_k$ 不可约而且 $\alpha_1+\cdots+\alpha_k=1$ 时,(3.11)成立等号的充分必要条件是存在正对角矩阵 $D_i\in\mathbb{R}^{n\times n}$ 和正数 γ_i, $i=2,\cdots,k$,使得

$$\gamma_i\left|A_i\right|=D_i^{-1}\left|A_1\right|D_i,\quad i=2,\cdots,k. \tag{3.12}$$

这是 H-矩阵对偶于 **7.3.8** 的结果,证明从略.

7.3.11 引理 设 $P=[p_{ij}]\in\mathbb{R}^{n\times n}$ 是不可约的非负矩阵,而且具有右 Perron 特征向量 $\boldsymbol{u}>0$ 和左 Perron 特征向量 $\boldsymbol{v}^{\mathrm{T}}>0$,则

$$\boldsymbol{u}^{\mathrm{T}}P\boldsymbol{v}\ge\boldsymbol{v}^{\mathrm{T}}P\boldsymbol{u}=\rho(P)\boldsymbol{v}^{\mathrm{T}}\boldsymbol{u}. \tag{3.13}$$

证 设 $\boldsymbol{u}=(u_1,\cdots,u_n)^{\mathrm{T}}$, $\boldsymbol{v}=(v_1,\cdots,v_n)^{\mathrm{T}}$,利用加权算术-几何平均不等式,有

$$\prod_{\substack{i,j=1\\p_{ij}>0}}^{n}\left(\frac{p_{ij}u_iv_j/(\boldsymbol{u}^{\mathrm{T}}P\boldsymbol{v})}{p_{ij}v_iu_j/(\boldsymbol{v}^{\mathrm{T}}P\boldsymbol{u})}\right)^{\frac{p_{ij}v_iu_j}{\boldsymbol{v}^{\mathrm{T}}P\boldsymbol{u}}}\le\sum_{\substack{i,j=1\\p_{ij}>0}}^{n}\left(\frac{p_{ij}u_iv_j/(\boldsymbol{u}^{\mathrm{T}}P\boldsymbol{v})}{p_{ij}v_iu_j/(\boldsymbol{v}^{\mathrm{T}}P\boldsymbol{u})}\right)\frac{p_{ij}v_iu_j}{\boldsymbol{v}^{\mathrm{T}}P\boldsymbol{u}}=1.$$

由此,并注意到 $Pu=\rho(P)u$ 和 $v^{\mathrm{T}}P=\rho(P)v^{\mathrm{T}}$,

$$\frac{u^{\mathrm{T}}Pv}{v^{\mathrm{T}}Pu}\geq\prod_{i,j=1}^{n}\left(\frac{u_i}{v_i}\right)^{\frac{p_{ij}v_iu_j}{v^{\mathrm{T}}Pu}}\prod_{i,j=1}^{n}\left(\frac{v_j}{u_j}\right)^{\frac{p_{ij}v_iu_j}{v^{\mathrm{T}}Pu}}$$

$$=\prod_{i=1}^{n}\left(\frac{u_i}{v_i}\right)^{\frac{\sum_{j=1}^{n}v_ip_{ij}u_j}{v^{\mathrm{T}}Pu}}\prod_{j=1}^{n}\left(\frac{v_j}{u_j}\right)^{\frac{\sum_{i=1}^{n}v_ip_{ij}u_j}{v^{\mathrm{T}}Pu}}$$

$$=\prod_{i=1}^{n}\left(\frac{u_i}{v_i}\right)^{\frac{v_iu_i}{v^{\mathrm{T}}u}}\prod_{j=1}^{n}\left(\frac{v_j}{u_j}\right)^{\frac{v_ju_j}{v^{\mathrm{T}}u}}=1.$$ □

7.3.12 定理 设 $A\in\mathbb{R}^{n\times n}$ 是 M-矩阵, 则

$$e^{\mathrm{T}}\left(A\circ A^{-1}\right)e\leq 1+e^{\mathrm{T}}\left(A_{ii}\circ A_{ii}^{-1}\right)e, \tag{3.14}$$

其中在不等号左边 $e\in\mathbb{R}^n$,在不等号右边 $e\in\mathbb{R}^{n-1}$,它们都是元素全为1的向量;而 A_{ii} 是在 A 中删去第 i 行和第 i 列后所得的 A 的 $(n-1)\times(n-1)$ 主子矩阵.

证 不失一般性,假定 $i=1$.于是(3.14)等价于

$$e^{\mathrm{T}}\left(A\circ A^{-1}-\begin{bmatrix}1 & 0\\ 0 & A_{11}\circ A_{11}^{-1}\end{bmatrix}\right)e\leq 0, \tag{3.15}$$

这里 $e\in\mathbb{R}^n$.令

$$B\equiv A-\frac{\det A}{\det A_{11}}E_{11},$$

其中 $E_{11}\in\mathbb{R}^{n\times n}$ 是(1,1)-位置的元素为1其余元素为0的矩阵.从行列式性质知

$$\det B=\det A-\frac{\det A}{\det A_{11}}\det A_{11}=0,$$

故 B 是奇异 M-矩阵.通过分块矩阵计算可得

$$A \circ A^{-1} - \begin{bmatrix} 1 & 0 \\ 0 & A_{11} \circ A_{11}^{-1} \end{bmatrix} = \frac{1}{\det A}(B \circ \mathrm{adj}B),$$

其中 adj$B$ 是 B 的转置伴随矩阵——转置余子式矩阵:

$$\mathrm{adj}B = \begin{bmatrix} (-1)^{1+1}\det B_{11} & (-1)^{2+1}\det B_{21} & \cdots & (-1)^{n+1}\det B_{n1} \\ (-1)^{1+2}\det B_{12} & (-1)^{2+2}\det B_{22} & \cdots & (-1)^{n+2}\det B_{n2} \\ \cdots & \cdots & \cdots & \cdots \\ (-1)^{1+n}\det B_{1n} & (-1)^{2+n}\det B_{2n} & \cdots & (-1)^{n+n}\det B_{nn} \end{bmatrix}.$$

B 有两条有用的性质:

(1) 若 $u, v \geq 0$ 是 $\mathbb{R}^n$ 中的非零向量,使得

$$Bu = 0,\ v^{\mathrm{T}}B = 0,$$

则 $u^{\mathrm{T}}Bv \leq 0$. 事实上,不失一般性,可以假定 B 不可约,而且 $u > 0, v > 0$. 于是,B 可表示为

$$B = \alpha I - P,\ P \geq 0,\ \alpha = \alpha(P),$$

其中 P 不可约.由此及 $Bu = 0$ 和 $v^{\mathrm{T}}B = 0$,推出

$$Pu = \alpha u,\ v^{\mathrm{T}}P = \alpha v^{\mathrm{T}},$$

即 u 和 v 分别是 P 的右和左 Perron 特征向量.从而

$$\begin{aligned} u^{\mathrm{T}}Bv &= \alpha u^{\mathrm{T}}v - u^{\mathrm{T}}Pv \\ &\leq \alpha u^{\mathrm{T}}v - v^{\mathrm{T}}Pu = \alpha u^{\mathrm{T}}v - \alpha u^{\mathrm{T}}v = 0, \end{aligned}$$

其中成立不等号应用了 **7.3.11**.

(2) $\mathrm{rank}B = n - 1$.因此,注意到

$$(\mathrm{adj}B)B = (\det B)I = 0,$$

可知 adj$B$ 是形如 cuv^{T} 的秩 1 矩阵,$c > 0$ 是常数,$u \geq 0$ 是 adj$B$ 的右 Perron 特征向量,而 $v \geq 0$ 是 adj$B$ 的左 Perron 特征向量.

现在可以完成定理的证明:

$$e^{\mathrm{T}}\left(A \circ A^{-1} - \begin{bmatrix} 1 & 0 \\ 0 & A_{11} \circ A_{11}^{-1} \end{bmatrix}\right)e$$

$$= \frac{1}{\det A} e^{\mathrm{T}} (B \circ \operatorname{adj} B) e = \frac{1}{\det A} e^{\mathrm{T}} \left(B \circ \left(cuv^{\mathrm{T}} \right) \right) e$$

$$= \frac{c}{\det A} u^{\mathrm{T}} B v \leq 0 .$$ □

7.3.13 定理 设 $A \in \mathbb{R}^{n \times n}$ 是 M-矩阵, 则

$$\operatorname{tr}\left(A^{-1} A^{\mathrm{T}}\right) = e^{\mathrm{T}}\left(A \circ A^{-1}\right) e \leq n, \tag{3.16}$$

其中 $e \in \mathbb{R}^n$ 是元素全为1的向量.等号成立的充分必要条件是 A 为对称矩阵.

证 逐次应用(3.14),对任何 M-矩阵 $A \in \mathbb{R}^{n \times n}$,以及任何指标集

$$\alpha = \{1,2,\cdots,n\} - \{i_1, i_2, \cdots, i_k\},\ 1 \leq i_1 < i_2 < \cdots < i_k \leq n,$$

成立

$$e^{\mathrm{T}}\left(A \circ A^{-1}\right) e \leq k + e^{\mathrm{T}}\left(A(\alpha) \circ A(\alpha)^{-1}\right) e .$$

(注意,这里不等号右端,$e \in \mathbb{R}^{n-k}$ 且元素全为1.)特别,

$$\begin{aligned} e^{\mathrm{T}}\left(A \circ A^{-1}\right) e &\leq 1 + e^{\mathrm{T}}\left(A(\{2,\cdots n\}) \circ A(\{2,\cdots,n\})^{-1}\right) e \\ &\leq 2 + e^{\mathrm{T}}\left(A(\{3,\cdots n\}) \circ A(\{3,\cdots,n\})^{-1}\right) e \\ &\leq \cdots \\ &\leq n - 1 + e^{\mathrm{T}}\left(A(\{n\}) \circ A(\{n\})^{-1}\right) e = n . \end{aligned}$$

由此并利用 **7.2.7**,即得(3.16).

A 为对称矩阵显然是(3.16)等号成立的充分条件,因为此时

$$\operatorname{tr}\left(A^{-1} A^{\mathrm{T}}\right) = \operatorname{tr}\left(A^{-1} A\right) = \operatorname{tr} I = n .$$

必要性的证明留作练习. □

7.3.14 定理 设 $A \in \mathbb{R}^{n \times n}$ 是 M-矩阵,则

$$\ell\left(A \circ A^{-1}\right) \leq 1, \tag{3.17}$$

这里 $\ell(A)$ 是 Z-矩阵 A 的最小特征值(见 **4.6.25**).而且,若 A 不可约,则(3.17)等号成立的充分必要条件是存在正对角矩阵 D 使得 AD

为对称矩阵.

证 对于可约的 A,矩阵 A 和 A^{-1},并因此 $A\circ A^{-1}$,有由同一指标集确定的不可约的组成部分.因为 $\ell\left(A\circ A^{-1}\right)$在其中某部分达到,所以只须对不可约的 A 证明不等式(3.17).

现在 $A\circ A^{-1}$ 是不可约的,依 **7.3.5**,它也是 M-矩阵.注意到

$$\ell\left(A\circ A^{-1}\right)\in\lambda\left(A\circ A^{-1}\right),$$

必有 $\ell\left(A\circ A^{-1}\right)>0$;而且

$$1/\ell\left(A\circ A^{-1}\right)$$

是 $\left(A\circ A^{-1}\right)^{-1}\geq 0$ 的谱半径.因此,$A\circ A^{-1}$ 存在相应于 $\ell\left(A\circ A^{-1}\right)$ 的正的特征向量 $x=\left(x_1,\cdots,x_n\right)^{\mathrm{T}}$,

$$\left(A\circ A^{-1}\right)x=\ell\left(A\circ A^{-1}\right)x,$$

取对角矩阵 $D=\operatorname{diag}\left(x_1,\cdots,x_n\right)$.因 $x=De$,$e=(1,\cdots,1)^{\mathrm{T}}\in\mathbb{R}^n$,故可得

$$\left[(AD)\circ(AD)^{-1}\right]e=\left[D^{-1}\left(A\circ A^{-1}\right)D\right]e=\ell\left(A\circ A^{-1}\right)e.$$

由此,并依 **7.2.7**,

$$\operatorname{tr}\left[(AD)^{-1}(AD)^{\mathrm{T}}\right]=e^{\mathrm{T}}\left[(AD)\circ(AD)^{-1}\right]e=n\ell\left(A\circ A^{-1}\right).$$

因为 AD 是 M-矩阵,依 **7.3.13** 有

$$n\ell\left(A\circ A^{-1}\right)=\operatorname{tr}\left[(AD)^{-1}(AD)^{\mathrm{T}}\right]\leq n,$$

所以(3.17)成立.

7.3.13 还蕴涵 $\ell\left(A\circ A^{-1}\right)=1$ 的充分必要条件是 AD 为对称矩阵. □

7.3.15 定理 设 $A,B\in\mathbb{R}^{n\times n}$ 是 M-矩阵,$B^{-1}=\left[\beta_{ij}\right]$,则

$$\ell\left(A\circ B^{-1}\right)\geq\ell(A)\min_{1\leq i\leq n}\beta_{ii},\tag{3.18}$$

这里 $\ell(A)$ 是 Z-矩阵 A 的最小特征值(见 **4.6.25**).

证 依 **4.6.5** 的(16),存在正对角矩阵 D,使得 $D^{-1}BD$ 是行对

角优势的.再注意到$(D^{-1}BD)^{-1}$和B^{-1}两者的对角元素是相同的.这样,

$$\ell(A\circ B^{-1})=\ell(D^{-1}(A\circ B^{-1})D)=\ell(A\circ(D^{-1}BD)^{-1}).$$

依 **2.2.6**,$(D^{-1}BD)^{-1}$是列对角优势的,从而有

$$A\circ(D^{-1}BD)^{-1}\geq A\mathrm{diag}(\beta_{11},\cdots,\beta_{nn}).$$

再依 **4.6.28** 和 **4.6.29**,得

$$\ell(A\circ B^{-1})\geq \ell(A\mathrm{diag}(\beta_{11},\cdots,\beta_{nn}))\geq \ell(A)\min_{1\leq i\leq n}\beta_{ii}.$$ □

7.3.16 推论 设$A\in\mathbb{R}^{n\times n}$是 M-矩阵,则

$$\ell(A\circ A^{-1})>\frac{1}{n}. \tag{3.19}$$

证 因为A^{-1}的对角元素是正的,所以可以选取正对角矩阵D,使得$(DA)^{-1}$的每个对角元素都是1.

因此

$$\mathrm{tr}((DA)^{-1})=n.$$

又因为DA是 M-矩阵,$(DA)^{-1}\geq 0$且$(DA)^{-1}$的每个特征值有正的实部,从而

$$\rho((DA)^{-1})<\mathrm{tr}((DA)^{-1}).$$

于是,依 **7.3.15**,

$$\ell(A\circ A^{-1})=\ell(D(A\circ A^{-1})D^{-1})=\ell((DA)\circ(DA)^{-1})$$

$$\geq \ell(DA)=\frac{1}{\rho((DA)^{-1})}>\frac{1}{\mathrm{tr}((DA)^{-1})}=\frac{1}{n}.$$ □

7.3.17 定理 设$A,B\in\mathbb{R}^{n\times n}$是 M-矩阵,$B$是不可约的.又设

$$u=(u_1,\cdots,u_n)^{\mathrm{T}}>0 \quad 和 \quad v=(v_1,\cdots,v_n)^{\mathrm{T}}>0$$

分别是B^{-1}的右 Perron 特征向量和左 Perron 特征向量.定义

$$w \equiv u \circ v = \left(w_1, \cdots, w_n\right)^{\mathrm{T}},$$

则

$$\ell\left(A \circ B^{-1}\right) \geq \frac{\ell(A)}{\ell(B)} \frac{\min\left\{w_1, \cdots, w_n\right\}}{\left(w_1 + \cdots + w_n\right)}, \tag{3.20}$$

这里 $\ell(A)$ 是 Z-矩阵 A 的最小特征值(见 **4.6.25**).

证 设 $B^{-1} \equiv \left[\beta_{ij}\right]$,令 $D = \operatorname{diag}\left(v_1, \cdots, v_n\right)$,由

$$v^{\mathrm{T}} B^{-1} = \rho\left(B^{-1}\right) v^{\mathrm{T}} = \ell(B)^{-1} v^{\mathrm{T}},$$

或写成

$$v^{\mathrm{T}} B = \ell(B) v^{\mathrm{T}},$$

推出 DB 是列对角优势 M-矩阵,依 **2.2.6**,$(DB)^{-1}$ 是行对角优势矩阵,得

$$\beta_{ij} \leq \beta_{ii} \frac{v_j}{v_i}, \quad i, j = 1, \cdots, n.$$

再利用 $B^{-1}u = \ell(B)^{-1} u$,有

$$\ell(B)^{-1} u_i = \sum_{j=1}^{n} \beta_{ij} u_j \leq \sum_{j=1}^{n} \beta_{ii} \frac{v_j u_j}{v_i}, \quad i = 1, \cdots, n.$$

于是

$$\frac{\min\left\{w_1, \cdots, w_n\right\}}{\ell(B) e^{\mathrm{T}} w} \leq \frac{w_i}{\ell(B) e^{\mathrm{T}} w} = \frac{u_i v_i}{\ell(B) u^{\mathrm{T}} v} \leq \beta_{ii}, \quad i = 1, \cdots, n,$$

其中 $e = (1, \cdots, 1)^{\mathrm{T}} \in \mathbb{R}^n$.由此,并利用 **7.3.15**,

$$\frac{\min\left\{w_1, \cdots, w_n\right\}}{\ell(B) e^{\mathrm{T}} w} \leq \min_{1 \leq i \leq n} \beta_{ii} \leq \frac{\ell\left(A \circ B^{-1}\right)}{\ell(A)}. \qquad \square$$

关于 Hadamard 积还有许多性质,在论著[HJ1991]中有比较详尽的叙述,而且还论及有关的若干应用.

7.4 单侧逆

7.4.1 引言 逆矩阵概念有多种推广,统称**广义逆**.这些广义逆保留了"逆"的若干重要属性,不仅对非奇异(方形)矩阵有定义,而且对奇异矩阵和长方阵也有定义.首先介绍矩阵单侧可逆性概念.

7.4.2 定义 矩阵 $A \in \mathbb{C}^{m\times n}$ 称为**左可逆**的,如果存在属于$\mathbb{C}^{n\times m}$ 的矩阵,记作 A_L^{-1},使得

$$A_L^{-1} A = I_n, \tag{4.1}$$

并称 A_L^{-1} 为 A 的**左逆**.

相仿,A 称为**右可逆**的,如果存在属于$\mathbb{C}^{n\times m}$ 的矩阵,记作 A_R^{-1},使得

$$AA_R^{-1} = I_m, \tag{4.2}$$

并称 A_R^{-1} 为 A 的**右逆**.

左逆和右逆统称**单侧逆**.

显然,若 $m = n$ 且 A 是非奇异的,则 A 的单侧逆和通常的逆 A^{-1} 是一致的.

下面将用到 **1.3.1** 和 **1.3.2** 中定义的线性映射的值域概念和零空间概念.在矩阵和线性映射一一对应的意义下,可以直接用矩阵来刻画值域和零空间.

设矩阵 $A \in \mathbb{C}^{m\times n}$,对应于 A 的**值域** $\mathscr{R}_A$ 是

$$\mathscr{R}_A = \{ Ax \in \mathbb{C}^m : x \in \mathbb{C}^n \}. \tag{4.3}$$

对应于 A 的**零空间** $\mathscr{N}_A$ 是

$$\mathscr{N}_A = \{x \in \mathbb{C}^n : Ax = 0\}. \tag{4.4}$$

7.4.3 定理　设 $A \in \mathbb{C}^{m\times n}$,下列条件等价:

(1) A 是左可逆的.

(2) $m \geq n$ 且 $\operatorname{rank}A = n$.

(3) A 的各列作为 $\mathbb{C}^m$ 中的向量是线性无关的.

(4) $\mathscr{N}_A = \{0\}$.

7.4.4 定理　设 $A \in \mathbb{C}^{m\times n}$,下列条件等价:

(1) A 是右可逆的.

(2) $m \leq n$ 且 $\operatorname{rank}A = m$.

(3) A 的各行作为 $\mathbb{C}^n$ 中的向量是线性无关的.

(4) $\mathscr{R}_A = \mathbb{C}^m$.

以上两定理证明比较容易,留作练习.

7.4.5 定理　设 $A \in \mathbb{C}^{m\times n}$ 是左可逆的,则 A 的所有左逆 A_L^{-1} 形如

$$A_L^{-1} = \begin{bmatrix} A_1^{-1} - BA_2A_1^{-1} & B \end{bmatrix} P, \tag{4.5}$$

其中 $P \in \mathbb{R}^{m\times m}$ 是转置矩阵,使得

$$PA = \begin{bmatrix} A_1 \\ A_2 \end{bmatrix}, \quad A_1 \in \mathbb{C}^{n\times n}, \quad \det A_1 \neq 0, \tag{4.6}$$

矩阵 $B \in \mathbb{C}^{n\times(m-n)}$ 是任意的.

证　A 左可逆.依 **7.4.3**, $m \geq n$, $\operatorname{rank}A = n$.因此方程

$$XA = I_n \tag{4.7}$$

是可解的.从 $\operatorname{rank}A = n$ 得知,确实存在 $m \times m$ 转置矩阵 P 使(4.6)成立.由于

$$XA = (XP)(PA),$$

方程(4.7)可以改写成

$$[Y_1 \quad Y_2]\begin{bmatrix} A_1 \\ A_2 \end{bmatrix} = I_n \quad 或写成 \quad Y_1 A_1 + Y_2 A_2 = I_n,$$

其中 $Y_1 \in \mathbb{C}^{n\times n}, Y_2 \in \mathbb{C}^{n\times(m-n)}$,且

$$[Y_1 \quad Y_2] = XP \quad 或等价地 \quad X = [Y_1 \quad Y_2]P.$$

于是,从 A_1 非奇异,得

$$Y_1 = A_1^{-1} - Y_2 A_2 A_1^{-1}.$$

因此,(4.7)的解即 A 的左逆

$$A_L^{-1} = X = [Y_1 \quad Y_2]P = [A_1^{-1} - Y_2 A_2 A_1^{-1} \quad Y_2]P,$$

其中 Y_2 可取任意的 $B \in \mathbb{C}^{n\times(m-n)}$. □

7.4.6 定理 设 $A \in \mathbb{C}^{m\times n}$ 是右可逆的,则 A 的所有右逆 A_R^{-1} 形如

$$A_R^{-1} = P\begin{bmatrix} A_1^{-1} - A_1^{-1} A_2 B \\ B \end{bmatrix}, \tag{4.8}$$

其中 $P \in \mathbb{R}^{n\times n}$ 是转置矩阵,使得

$$AP = [A_1 \quad A_2], \quad A_1 \in \mathbb{C}^{m\times m}, \quad \det A_1 \neq 0. \tag{4.9}$$

矩阵 $B \in \mathbb{C}^{(n-m)\times m}$ 是任意的.

证明与 **7.4.5** 相仿,留作练习.

下面考虑方程

$$Ax = b, \ A \in \mathbb{C}^{m\times n}, \ x \in \mathbb{C}^n, \ b \in \mathbb{C}^m. \tag{4.10}$$

7.4.7 定理 设 $A \in \mathbb{C}^{m\times n}$ 是左可逆的,则方程(4.10)可解的充分必要条件是

$$(I_m - AA_L^{-1})b = 0. \tag{4.11}$$

而且,若(4.11)成立,则(4.10)的解还是唯一的,并为

$$x = A_L^{-1} b. \tag{4.12}$$

证 设 x 满足(4.10),则

$$AA_L^{-1}b = A\left(A_L^{-1}A\right)x = Ax = b,$$

推出(4.11).反之,若(4.11)成立,则向量 $x = A_L^{-1}b$ 是(4.10)的解.如果 $\tilde{x}$ 也是(4.10)的解,那么 $A(x-\tilde{x})=0$,从而由 **7.4.3** 的(4)推出 $\tilde{x}=x$. □

7.4.8 定理 设 $A\in\mathbb{C}^{m\times n}$ 是右可逆的,则对任何 $b\in\mathbb{C}^m, b\neq 0$,方程(4.10)是可解的,每个解 $x\in\mathbb{C}^n$ 形如

$$x = A_R^{-1}b. \tag{4.13}$$

证 方程(4.10)的可解性是明显的,因为

$$A\left(A_R^{-1}b\right)=b,$$

从而形如(4.13)的 x 是(4.10)的解.

现在假设 x 是(4.10)的解,如同(4.9),将 A 分裂成

$$A=\begin{bmatrix}A_1 & A_2\end{bmatrix}P,\ \det A_1\neq 0,$$

其中 $P\in\mathbb{C}^{n\times n}$ 是置换矩阵.而且对 x 作相容的分裂:

$$x = P\begin{bmatrix}x^{(1)}\\ x^{(2)}\end{bmatrix},$$

则方程(4.10)可改写成

$$A_1x^{(1)}+A_2x^{(2)}=b,$$

因此

$$x^{(1)}=A_1^{-1}b-A_1^{-1}A_2x^{(2)}.$$

利用(4.8),并注意到 $b\neq 0$,必存在 $C\in\mathbb{C}^{(n-m)\times m}$ 使得

$$Cb=x^{(2)},$$

推出

$$A_R^{-1}b=P\begin{bmatrix}A_1^{-1}b-A_1^{-1}A_2Cb\\ Cb\end{bmatrix}=P\begin{bmatrix}x^{(1)}\\ x^{(2)}\end{bmatrix}=x. \qquad \square$$

7.4.9 定理 设 $A\in\mathbb{C}^{m\times n}, b\in\mathbb{C}^m, X\in\mathbb{C}^{n\times m}$ 是方程

$$AXA=A \tag{4.14}$$

的解.则方程(4.10)有解的充分必要条件是

$$AXb = b. \tag{4.15}$$

而且,当(4.10)有解时,其任何解形如

$$x = Xb + (I - XA)y, \quad y \in \mathbb{C}^n. \tag{4.16}$$

证 定理所假设的条件(4.14)等价于

$$0 = (A - AXA)y = A(I - XA)y, \quad \forall\ y \in \mathbb{C}^n.$$

这表明方程 $Ax = 0$ 的通解为

$$x = (I - XA)y, \quad \forall\ y \in \mathbb{C}^n. \tag{4.17}$$

现在假设(4.10)有解 x, $Ax = b$,则由(4.14)有

$$AXb = AX(Ax) = (AXA)x = Ax = b.$$

反之,若成立(4.15),则

$$x = Xb, \tag{4.18}$$

就是(4.10)的一个解.

当(4.10)有解时,联合(4.17)和(4.18),即得(4.16). □

7.4.10 定理 设 $A \in \mathbb{C}^{m\times n}$ 是左可逆的,则

(1) AA_L^{-1} 是等幂矩阵.

(2) AA_L^{-1} 是$\mathbb{C}^m \to \mathscr{R}_A$ 上的投影算子.

(3) $\dim \mathscr{N}_{A_L^{-1}} = m - \dim \mathscr{R}_A$.

证 (1) 直接计算,

$$\left(AA_L^{-1}\right)^2 = \left(AA_L^{-1}\right)\left(AA_L^{-1}\right) = A\left(A_L^{-1}A\right)A_L^{-1} = AA_L^{-1}.$$

(2) 根据(1)及 **1.3.13**, AA_L^{-1} 是投影算子;而且,对于任一 $x \in \mathbb{C}^m$,显然

$$\left(AA_L^{-1}\right)x = A\left(A_L^{-1}x\right) \in \mathscr{R}_A.$$

剩下要证明 AA_L^{-1} 是从$\mathbb{C}^m$ 到 $\mathscr{R}_A$ 的满映射.事实上,若 $x \in \mathscr{R}_A$,则存在 $y \in \mathbb{C}^n$, $x = Ay$,于是

$$\left(AA_L^{-1}\right)x = \left(AA_L^{-1}\right)Ay = A\left(A_L^{-1}A\right)y = Ay = x \in \mathscr{R}_A.$$

(3) 依 **1.3.3**,

$$\dim \mathscr{N}_{AA_L^{-1}} + \dim \mathscr{R}_{AA_L^{-1}} = \dim \mathbb{C}^m = m .$$

利用(2), AA_L^{-1} 是$\mathbb{C}^m \to \mathscr{R}_A$ 的满映射,表明

$$\mathscr{R}_{AA_L^{-1}} = \mathscr{R}_A .$$

另一方面,显然 $\mathscr{N}_{A_L^{-1}} \subset \mathscr{N}_{AA_L^{-1}}$;反之,若 $x \in \mathscr{N}_{AA_L^{-1}}$,则从 $AA_L^{-1}x = 0$ 得

$$A_L^{-1}x = A_L^{-1}\left(AA_L^{-1}x\right) = 0 ,$$

推出 $\mathscr{N}_{AA_L^{-1}} \subset \mathscr{N}_{A_L^{-1}}$.因此

$$\mathscr{N}_{AA_L^{-1}} = \mathscr{N}_{A_L^{-1}} .$$ □

7.4.11 定理 设 $A \in \mathbb{C}^{m\times n}$ 是右可逆的,则

(1) $A_R^{-1}A$ 是等幂矩阵.

(2) $I - A_R^{-1}A$ 是$\mathbb{C}^n \to \mathscr{N}_A$ 上的投影算子.

(3) $\dim \mathscr{R}_{A_L^{-1}} = n - \dim \mathscr{N}_A$.

证明与 **7.4.10** 相仿,留作练习.

7.5 广义逆 A^{I}

7.5.1 引言 对于 $A \in \mathbb{C}^{m\times n}$ 的单侧逆概念而言,**7.4.3** 和 **7.4.4** 表明,当且仅当 A 为满秩矩阵时单侧逆才是存在的、有意义的.矩阵 $A \in \mathbb{C}^{m\times n}$ **满秩**是指

$$\operatorname{rank}A = \min\{m, n\}. \tag{5.1}$$

进一步的目的是推广单侧逆的概念,使之适用于任何矩阵 $A \in \mathbb{C}^{m\times n}$.注意到 A 如若存在单侧逆 $X \in \mathbb{C}^{n\times m}$ 的话, X 无论是左

逆或右逆,都应满足方程

$$AXA = A, \quad XAX = X.$$

由此出发,可以建立一般的广义逆概念.

7.5.2 定义 设 $A \in \mathbb{C}^{m\times n}$,矩阵 $A^{\mathrm{I}} \in \mathbb{C}^{n\times m}$ 称为 A 的**广义逆**,如果成立

$$AA^{\mathrm{I}}A = A, \tag{5.2}$$

$$A^{\mathrm{I}}AA^{\mathrm{I}} = A^{\mathrm{I}}. \tag{5.3}$$

7.5.3 定理 对于任何 $A \in \mathbb{C}^{m\times n}$,存在广义逆 $A^{\mathrm{I}} \in \mathbb{C}^{n\times m}$.

证 首先,$m\times n$ 零矩阵显然以 $n\times m$ 零矩阵为其广义逆.

现设 $\mathrm{rank}A = r \neq 0$,则依 **9.6.9**,存在非奇异矩阵 $R \in \mathbb{C}^{m\times m}$ 和 $V \in \mathbb{C}^{n\times n}$,使得

$$A = R\begin{bmatrix} I_r & 0 \\ 0 & 0 \end{bmatrix}V. \tag{5.4}$$

容易验证,形如

$$A^{\mathrm{I}} = V^{-1}\begin{bmatrix} I_r & B_1 \\ B_2 & B_2B_1 \end{bmatrix}R^{-1},$$

$$B_1 \in \mathbb{C}^{r\times(m-r)}, \quad B_2 \in \mathbb{C}^{(n-r)\times r} \tag{5.5}$$

的矩阵 A^{I},满足条件(5.2)和(5.3). □

因此,广义逆总是存在的,但一般不是唯一的.

7.5.4 定理 $A \in \mathbb{C}^{m\times n}$ 有唯一广义逆当且仅当 $A = 0$ 或 A 为非奇异方阵.

利用分解形式(5.4)容易推出这一结论,留作练习.

7.5.5 定理 设 $A \in \mathbb{C}^{m\times n}$, A^{I} 是 A 的一个广义逆,则

(1) $A^{I}A$ 和 AA^{I} 是等幂矩阵.

(2) $\operatorname{rank}A^{I}=\operatorname{rank}A$.

(3) $\left(A^{I}\right)^{*}$ 是 A^{*} 的一个广义逆.

证 依 **7.5.2**,

$$AA^{I}A=A,\quad A^{I}AA^{I}=A^{I}.$$

因此,有

$$\left(A^{I}A\right)^{2}=\left(A^{I}A\right)\left(A^{I}A\right)=\left(A^{I}AA^{I}\right)A=A^{I}A,$$
$$\left(AA^{I}\right)^{2}=\left(AA^{I}\right)\left(AA^{I}\right)=\left(AA^{I}A\right)A^{I}=AA^{I}$$

和

$$A^{*}=\left(AA^{I}A\right)^{*}=A^{*}\left(A^{I}\right)^{*}A^{*},$$
$$\left(A^{I}\right)^{*}=\left(A^{I}AA^{I}\right)^{*}=\left(A^{I}\right)^{*}A^{*}\left(A^{I}\right)^{*}$$

从而(1)和(3)得证.而且

$$\operatorname{rank}A=\operatorname{rank}\left(AA^{I}A\right)\leq\min\left\{\operatorname{rank}A,\operatorname{rank}A^{I}\right\},$$
$$\operatorname{rank}A^{I}=\operatorname{rank}\left(A^{I}AA^{I}\right)\leq\min\left\{\operatorname{rank}A^{I},\operatorname{rank}A\right\}$$

推出(2)成立. □

最后的证明利用了

$$\operatorname{rank}(AB)\leq\min\{\operatorname{rank}A,\operatorname{rank}B\},\quad \forall A\in\mathbb{C}^{m\times n},B\in\mathbb{C}^{n\times l}. \tag{5.6}$$

这一不等式可以直接从如下事实推出: AB 的每一行是 B 的各行的线性组合,其每一列是 A 的各列的线性组合.

7.5.6 定理 设 $A\in\mathbb{C}^{m\times n}$, A^{I} 是 A 的一个广义逆,则

$$\mathscr{R}_{A}\oplus\mathscr{N}_{A^{I}}=\mathbb{C}^{m}, \tag{5.7}$$

$$\mathscr{N}_{A}\oplus\mathscr{R}_{A^{I}}=\mathbb{C}^{n}. \tag{5.8}$$

证 依 **7.5.5** 的(1), AA^{I} 是投影算子.显然, $\mathscr{R}_{AA^{I}}\subset\mathscr{R}_{A}$;反之,若 $y=Ax\in\mathscr{R}_{A}$,则

$$y = AA^{\mathrm{I}}Ax = (AA^{\mathrm{I}})(Ax) \in \mathscr{R}_{AA^{\mathrm{I}}},$$

推出 $\mathscr{R}_A \subset \mathscr{R}_{AA^{\mathrm{I}}}$.因而,$\mathscr{R}_{AA^{\mathrm{I}}} = \mathscr{R}_A$,即 AA^{I} 是$\mathbb{C}^m \to \mathscr{R}_A$ 上的映射.

由此及 **7.5.5** 的(2),有

$$\mathrm{rank}A = \mathrm{rank}(AA^{\mathrm{I}}) = \mathrm{rank}A^{\mathrm{I}}.$$

于是,因为显然有 $\mathscr{N}_{A^{\mathrm{I}}} \subset \mathscr{N}_{AA^{\mathrm{I}}}$,而且依 **1.3.3**,

$$\dim \mathscr{N}_{A^{\mathrm{I}}} = m - \mathrm{rank}A^{\mathrm{I}} = m - \mathrm{rank}(AA^{\mathrm{I}}) = \dim \mathscr{N}_{AA^{\mathrm{I}}},$$

推出 $\mathscr{N}_{AA^{\mathrm{I}}} = \mathscr{N}_{A^{\mathrm{I}}}$.

联合以上结果及 AA^{I} 是投影算子这一事实,得

$$\mathbb{C}^m = \mathscr{R}_{AA^{\mathrm{I}}} \oplus \mathscr{N}_{AA^{\mathrm{I}}} = \mathscr{R}_A \oplus \mathscr{N}_{A^{\mathrm{I}}},$$

(5.7)得证.

由于定义广义逆 A^{I} 的方程中 A 和 A^{I} 的地位是对称的,在(5.7)中分别用 A^{I} 和 A 替换 A 和 A^{I},并相应地用$\mathbb{C}^n$ 替换$\mathbb{C}^m$,即得(5.8). □

如果 $A \in \mathbb{C}^{n\times n}$ 是非奇异矩阵,显然成立

$$\mathscr{R}_A \oplus \mathscr{N}_{A^{\mathrm{I}}} = \mathbb{C}^n \oplus \{0\} = \mathbb{C}^n,$$

$$\mathscr{N}_A \oplus \mathscr{R}_{A^{\mathrm{I}}} = \{0\} \oplus \mathbb{C}^n = \mathbb{C}^n.$$

上述定理是将这些结果推广到任意矩阵及其广义逆.

7.5.7 推论 设 $A \in \mathbb{C}^{m\times n}$,$A^{\mathrm{I}}$ 是 A 的一个广义逆,则

(1) AA^{I} 是$\mathbb{C}^m \to \mathscr{R}_A$ 上的投影算子.

(2) $A^{\mathrm{I}}A$ 是$\mathbb{C}^n \to \mathscr{R}_{A^{\mathrm{I}}}$ 上的投影算子.

(3) $AA^{\mathrm{I}}x = x, \forall x \in \mathscr{R}_A$.

(4) $A^{\mathrm{I}}Ax = x, \forall x \in \mathscr{R}_{A^{\mathrm{I}}}$.

证明留作练习.提示:(1)和(2)的证明可仿照 **7.4.10**.(3)和(4)是(1)和(2)的直接推论.

7.5.8 定理 设 $A \in \mathbb{C}^{m\times n}$,并设 $\Phi \in \mathbb{C}^m$ 和 $\Psi \in \mathbb{C}^n$ 是子空间,使得

$$\mathscr{R}_A \oplus \Phi = \mathbb{C}^m, \quad \mathscr{N}_A \oplus \Psi = \mathbb{C}^n, \tag{5.9}$$

则存在 A 的广义逆 A^{I},使得

$$\mathscr{N}_{A^{\mathrm{I}}} = \Phi, \quad \mathscr{R}_{A^{\mathrm{I}}} = \Psi. \tag{5.10}$$

证 设 $\operatorname{rank}A = r$,并将 A 写成(5.4)的形式:

$$A = RJV, \quad J = \begin{bmatrix} I_r & 0 \\ 0 & 0 \end{bmatrix},$$

于是

$$J(Vy) = R^{-1}(Ay) = 0, \quad \forall y \in \mathscr{N}_A,$$

推出

$$Vy \in \mathscr{N}_J, \quad \forall y \in \mathscr{N}_A.$$

类似地,

$$R^{-1}x = R^{-1}(Ay) = J(Vy), \quad \forall x = Ay \in \mathscr{R}_A,$$

推出

$$R^{-1}x \in \mathscr{R}_J, \quad \forall x \in \mathscr{R}_A.$$

因此,若令

$$\hat{\Phi} = \{R^{-1}x : x \in \Phi\}, \quad \hat{\Psi} = \{Vy : y \in \Psi\},$$

则从(5.9)有

$$\mathscr{R}_J \oplus \hat{\Phi} = \mathbb{C}^m, \quad \mathscr{N}_J \oplus \hat{\Psi} = \mathbb{C}^n.$$

由此可知,存在矩阵 $X \in \mathbb{C}^{r\times(m-r)}$ 和 $Y \in \mathbb{C}^{(n-r)\times r}$,成立

$$\Phi = \left\{ Rx : x = \begin{bmatrix} X \\ I_{m-r} \end{bmatrix} z, \forall z = (z_1, \cdots, z_{m-r}) \right\}$$

和

$$\Psi = \left\{ V^{-1}y : y = \begin{bmatrix} I_r \\ Y \end{bmatrix} z, \forall z = (z_1, \cdots, z_r) \right\}.$$

现在定义 A^{I} 为

$$A^{\mathrm{I}} = V^{-1}\begin{bmatrix} I_r & -X \\ Y & -YX \end{bmatrix} R^{-1},$$

容易验证 A^{I} 是 A 的广义逆,而且具有性质(5.10). □

这一定理是 **7.5.6** 的逆定理.

7.5.9 定理 设 $A \in \mathbb{C}^{m\times n}$, $(A^* A)^{\mathrm{I}}$ 是的任一广义逆,则

$$P = A(A^* A)^{\mathrm{I}} A^*$$

是$\mathbb{C}^m \to \mathcal{R}_A$ 上的正交投影算子.

证 根据(5.2),

$$A^* A(A^* A)^{\mathrm{I}} A^* A = A^* A.$$

再依 **1.7.20** 的(6)和(7),

$$A(A^* A)^{\mathrm{I}} A^* A = A, \ A^* A(A^* A)^{\mathrm{I}} A^* = A^*. \qquad (5.11)$$

这两个关系式都可直接推出 $P^2 = P$.

另一方面,由于

$$P^* = A\left(\left(A^* A\right)^{\mathrm{I}}\right)^* A^* = A\left(A^* A\right)^{\mathrm{J}} A^*, \qquad (5.12)$$

其中

$$\left(A^* A\right)^{\mathrm{J}} \equiv \left(\left(A^* A\right)^{\mathrm{I}}\right)^*.$$

从 **7.5.5** 的(3)得知$(A^* A)^{\mathrm{J}}$也是 $A^* A$ 的一个广义逆.于是,对$(A^* A)^{\mathrm{J}}$也成立与(5.11)相仿的关系式,而且利用这些关系式可得

$$(P-P^*)(P-P^*)^* = PP^* + P^*P - P^2 - (P^*)^2 = 0,$$

因此,$P^* = P$.

另外,显然,$Px \in \mathcal{R}_A, \forall x \in \mathbb{C}^m$.

至此,已证明 P 是$\mathbb{C}^m$ 到 $\mathcal{R}_A$ 的正交投影算子.

为了证明 P 还是$\mathbb{C}^m \to \mathcal{R}_A$ 的满映射.设 $y \in \mathcal{R}_A$,则存在

$x \in \mathbb{C}^n$ 使得 $y = Ax$,利用(5.11)中的第一等式,得

$$y = A\left(A^* A\right)^{I} A^*\left(Ax\right) = Py ,$$

注意,$y \in \mathscr{R}_A \subset \mathbb{C}^m$. □

7.6 Moore-Penrose 逆

7.6.1 引言 如果 A^{I} 是 $A \in \mathbb{C}^{m\times n}$ 的广义逆,从 **7.5.7** 知道,当 AA^{I} 和 $A^{I}A$ 是自伴矩阵时,AA^{I} 是$\mathbb{C}^m \to \mathscr{R}_A$ 上的正交投影算子, $A^{I}A$ 是 $\mathbb{C}^n \to \mathscr{R}_{A^{I}}$ 上的正交投影算子.如此,广义逆将可引出非常良好的性质.于是,自然地对广义逆的概念增添 AA^{I} 和 $A^{I}A$ 为自伴的条件:

$$\left(AA^{I}\right)^* = AA^{I}, \quad \left(A^{I}A\right)^* = A^{I}A .$$

7.6.2 定义 设 $A \in \mathbb{C}^{m\times n}$,矩阵 $A^+ \in \mathbb{C}^{n\times m}$ 称为矩阵 A 的 **Moore-Penrose(广义)逆**,如果成立

$$AA^+ A = A , \tag{6.1}$$

$$A^+ AA^+ = A^+ , \tag{6.2}$$

$$\left(AA^+\right)^* = AA^+ , \tag{6.3}$$

$$\left(A^+ A\right)^* = A^+ A . \tag{6.4}$$

7.6.3 定理 设 $A \in \mathbb{C}^{m\times n}$,矩阵 $A^{I} \in \mathbb{C}^{n\times m}$ 是 A 的广义逆,则 A^{I} 为 A 的 Moore-Penrose 逆的充分必要条件是

$$\mathscr{N}_{A^{I}} = \left(\mathscr{R}_A\right)^{\perp}, \quad \mathscr{R}_{A^{I}} = \left(\mathscr{N}_A\right)^{\perp} . \tag{6.5}$$

也就是说,$\mathscr{N}_{A^{I}}$ 和 $\mathscr{R}_A$ 互为正交补,$\mathscr{R}_{A^{I}}$ 和 $\mathscr{N}_A$ 互为正交补.

证 A^{I} 是 A 的广义逆,依 **7.5.6**,成立

$$\mathscr{R}_A \oplus \mathscr{N}_{A^\mathrm{I}} = \mathbb{C}^m, \tag{6.6}$$

$$\mathscr{N}_A \oplus \mathscr{R}_{A^\mathrm{I}} = \mathbb{C}^n. \tag{6.7}$$

因此,依 **7.6.2**,对于本定理,要证明的是: AA^I 和 $A^\mathrm{I}A$ 自伴,即

$$\left(AA^\mathrm{I}\right)^* = AA^\mathrm{I}, \quad \left(A^\mathrm{I}A\right)^* = A^\mathrm{I}A, \tag{6.8}$$

等价于 $\mathscr{R}_A$ 和 $\mathscr{N}_{A^\mathrm{I}}$ 正交,$\mathscr{N}_A$ 和 $\mathscr{R}_{A^\mathrm{I}}$ 正交,即

$$\begin{aligned} &(x,y)=0, \quad \forall x \in \mathscr{R}_A, y \in \mathscr{N}_{A^\mathrm{I}}; \\ &(x,y)=0, \quad \forall x \in \mathscr{N}_A, y \in \mathscr{R}_{A^\mathrm{I}}. \end{aligned} \tag{6.9}$$

现在设(6.8)成立,则对任意的 $x \in \mathscr{R}_A$, $y \in \mathscr{N}_{A^\mathrm{I}}$,由于对 x 存在 $z \in \mathbb{C}^n$ 使得 $x = Az$,对 y 满足 $A^\mathrm{I}y = 0$,并利用

$$\left(AA^\mathrm{I}\right)^* = AA^\mathrm{I}, \quad AA^\mathrm{I}A = A,$$

有

$$\begin{aligned} (x,y) &= (Az,y) = \left(AA^\mathrm{I}Az, y\right) \\ &= \left(Az, \left(AA^\mathrm{I}\right)^* y\right) = \left(x, AA^\mathrm{I}y\right) = (x,0) = 0. \end{aligned}$$

类似地,对任意的 $x \in \mathscr{N}_A$, $y \in \mathscr{R}_{A^\mathrm{I}}$,由于对 x 满足 $Ax = 0$,对 y 存在 $z \in \mathbb{C}^m$ 使得 $y = A^\mathrm{I}z$,并利用

$$\left(A^\mathrm{I}A\right)^* = A^\mathrm{I}A, \quad A^\mathrm{I}AA^\mathrm{I} = A^\mathrm{I},$$

有

$$\begin{aligned} (x,y) &= \left(x, A^\mathrm{I}z\right) = \left(x, A^\mathrm{I}AA^\mathrm{I}z\right) \\ &= \left(\left(A^\mathrm{I}A\right)^* x, A^\mathrm{I}z\right) = \left(A^\mathrm{I}Ax, y\right) = (0,y) = 0. \end{aligned}$$

反之,设(6.9)成立,则对任意的

$$u = x' + y', v = x'' + y'' \in \mathbb{C}^m,$$
$$x', x'' \in \mathscr{R}_A, \quad y', y'' \in \mathscr{N}_{A^\mathrm{I}},$$

由于对 x', x'' 存在 $z', z'' \in \mathbb{C}^n$ 使得

$$x' = Az', \quad x'' = Az''.$$

对 y', y'' 满足

$$A^{\mathrm{I}}y' = A^{\mathrm{I}}y'' = 0.$$

并利用 $AA^{\mathrm{I}}A = A$ 和(6.9),有

$$\begin{aligned}\left(\left(\left(AA^{\mathrm{I}}\right)^* - AA^{\mathrm{I}}\right)u, v\right) &= \left(\left(AA^{\mathrm{I}}\right)^* u, v\right) - \left(AA^{\mathrm{I}}u, v\right) \\ &= \left(u, AA^{\mathrm{I}}v\right) - \left(AA^{\mathrm{I}}u, v\right) \\ &= \left(x' + y', AA^{\mathrm{I}}x'' + AA^{\mathrm{I}}y''\right) - \left(AA^{\mathrm{I}}x' + AA^{\mathrm{I}}y', x'' + y''\right) \\ &= \left(x' + y', AA^{\mathrm{I}}Az''\right) - \left(AA^{\mathrm{I}}Az', x'' + y''\right) \\ &= \left(x' + y', x''\right) - \left(x', x'' + y''\right) \\ &= \left(y', x''\right) - \left(x', y''\right) = 0 - 0 = 0,\end{aligned}$$

表明 $\left(AA^{\mathrm{I}}\right)^* - AA^{\mathrm{I}} = 0$.

类似地可以证明 $\left(A^{\mathrm{I}}A\right)^* - A^{\mathrm{I}}A = 0$. □

7.6.4 引理 对于任何矩阵 $A \in \mathbf{C}^{m\times n}$,成立

$$\mathscr{N}_{A^*} = \left(\mathscr{R}_A\right)^{\perp}, \quad \mathscr{R}_A \oplus \mathscr{N}_{A^*} = \mathbf{C}^m, \tag{6.10}$$

$$\mathscr{R}_{A^*} = \left(\mathscr{N}_A\right)^{\perp}, \quad \mathscr{N}_A \oplus \mathscr{R}_{A^*} = \mathbf{C}^n. \tag{6.11}$$

证 一方面,对于任意的 $x \in \mathscr{R}_A, y \in \mathscr{N}_{A^*}$,由于对 x 存在 $z \in \mathbf{C}^n$ 使得 $x = Az$,对 y 满足 $A^*y = 0$,有

$$(x, y) = (Az, y) = \left(z, A^*y\right) = (z, 0) = 0,$$

表明 $\mathscr{R}_A$ 和 $\mathscr{N}_{A^*}$ 正交.

另一方面,注意到 A 和 A^* 有相同的秩.设

$$\operatorname{rank}A = \operatorname{rank}A^* = r,$$

则有 $\dim \mathscr{R}_A = r$;而 $A^*x = 0$ 存在 $m - r$ 个线性无关的解,故有 $\dim \mathscr{N}_{A^*} = m - r$.

综合以上两方面,$\mathscr{R}_A \oplus \mathscr{N}_{A^*}$ 的维数是

$$\dim \mathscr{R}_A + \dim \mathscr{N}_{A^*} = m,$$

因此,(6.10)成立.

类似地,可以证明(6.11)也成立. □

7.6.5 定理 对于任何矩阵 $A\in\mathbb{C}^{m\times n}$ 存在唯一的 Moore-Penrose 逆 $A^{+}\in\mathbb{C}^{n\times m}$.

证 子空间的正交补的存在性,以及 **7.5.8** 和 **7.6.3**,保证了 A^{+} 的存在性.下面证明其唯一性.

假定 A 还有一个 Moore-Penrose 逆 $\hat{A}^{+}$,则根据条件(6.2)和(6.3),有

$$\left(A^{+}\right)^{*}=\left(A^{+}\right)^{*}A^{*}\left(A^{+}\right)^{*}=\left(AA^{+}\right)^{*}\left(A^{+}\right)^{*}=AA^{+}\left(A^{+}\right)^{*}$$

和

$$\left(\hat{A}^{+}\right)^{*}=\left(\hat{A}^{+}\right)^{*}A^{*}\left(\hat{A}^{+}\right)^{*}=\left(A\hat{A}^{+}\right)^{*}\left(\hat{A}^{+}\right)^{*}=A\hat{A}^{+}\left(\hat{A}^{+}\right)^{*},$$

于是

$$\left(\hat{A}^{+}\right)^{*}-\left(A^{+}\right)^{*}=A\left(\hat{A}^{+}\left(\hat{A}^{+}\right)^{*}-A^{+}\left(A^{+}\right)^{*}\right),$$

从而

$$\mathscr{R}_{\left(\hat{A}^{+}\right)^{*}-\left(A^{+}\right)^{*}}\subset\mathscr{R}_{A}. \tag{6.12}$$

另一方面,利用(6.1)和(6.4),有

$$A^{*}=A^{*}\left(A^{+}\right)^{*}A^{*}=\left(A^{+}A\right)^{*}A^{*}=A^{+}AA^{*}$$

和

$$A^{*}=A^{*}\left(\hat{A}^{+}\right)^{*}A^{*}=\left(\hat{A}^{+}A\right)^{*}A^{*}=\hat{A}^{+}AA^{*},$$

于是

$$\left(\hat{A}^{+}-A^{+}\right)AA^{*}=0\quad\text{或}\quad AA^{*}\left(\left(\hat{A}^{+}\right)^{*}-\left(A^{+}\right)^{*}\right)=0.$$

由此并依 **1.7.20** 的(5),得

$$\mathscr{R}_{\left(\hat{A}^{+}\right)^{*}-\left(A^{+}\right)^{*}}\subset\mathscr{N}_{AA^{*}}=\mathscr{N}_{A^{*}}. \tag{6.13}$$

依 **7.6.4**, $\mathscr{N}_{A^{*}}$ 正交于 $\mathscr{R}_{A}$.因此,从(6.12)和(6.13)推出

$$\left(\hat{A}^{+}\right)^{*}-\left(A^{+}\right)^{*}=0\quad\text{即}\quad\hat{A}^{+}=A^{+}.$$ □

7.6.6 定理　对于任何矩阵 $A\in\mathbb{C}^{m\times n}$,成立

$$\mathscr{N}_{A^+}=\mathscr{N}_{A^*},\quad \mathscr{R}_{A^+}=\mathscr{R}_{A^*}. \tag{6.14}$$

证　比较(6.10)和(6.6),(6.11)和(6.7),即得(6.14).　□

7.6.7 定理　设 $A\in\mathbb{C}^{m\times n}$, $\operatorname{rank}A=r$,并设

$$A=FR^*,\quad F\in\mathbb{C}^{m\times r},\ R^*\in\mathbb{C}^{r\times n} \tag{6.15}$$

是 A 的一个**秩分解**,即满足 $\operatorname{rank}F=\operatorname{rank}R^*=r$,则

$$A^+=R\left(R^*R\right)^{-1}\left(F^*F\right)^{-1}F^*. \tag{6.16}$$

特别,如果 A 是满秩矩阵,即 $\operatorname{rank}A=\min\{m,n\}$,那么

$$A^+=\begin{cases}A^*\left(AA^*\right)^{-1}, & m\le n,\\ \left(A^*A\right)^{-1}A^*, & m\ge n.\end{cases} \tag{6.17}$$

注意:当 $m=n$ 时表示式(6.17)成为

$$A^+=A^{-1}.$$

证　注意到 F^*F 和 R^*R 是 $r\times r$ 的满秩矩阵,因而是可逆的.利用(6.1),有

$$\begin{aligned}R^*A^+F&=\left(F^*F\right)^{-1}\left(F^*F\right)R^*A^+F\left(R^*R\right)\left(R^*R\right)^{-1}\\&=\left(F^*F\right)^{-1}F^*AA^+AR\left(R^*R\right)^{-1}\\&=\left(F^*F\right)^{-1}F^*AR\left(R^*R\right)^{-1}\\&=\left(F^*F\right)^{-1}F^*FR^*R\left(R^*R\right)^{-1}=I_r.\end{aligned} \tag{6.18}$$

利用(6.3)及(6.18),有

$$\begin{aligned}R^*A^+&=\left(F^*F\right)^{-1}\left(F^*F\right)R^*A^+=\left(F^*F\right)^{-1}F^*AA^+\\&=\left(F^*F\right)^{-1}F^*\left(AA^+\right)^*=\left(F^*F\right)^{-1}F^*\left(A^+\right)^*\left(RF^*\right)\\&=\left(F^*F\right)^{-1}\left(R^*A^+F\right)^*F^*=\left(F^*F\right)^{-1}F^*.\end{aligned} \tag{6.19}$$

利用(6.4)及(6.18),有

$$A^+F=A^+F\left(R^*R\right)\left(R^*R\right)^{-1}=A^+AR\left(R^*R\right)^{-1}$$

$$= \left(A^+A\right)^* R\left(R^*R\right)^{-1} = \left(RF^*\right)\left(A^+\right)^* R\left(R^*R\right)^{-1}$$
$$= R\left(R^*A^+F\right)^*\left(R^*R\right)^{-1} = R\left(R^*R\right)^{-1}. \tag{6.20}$$

最后,利用(6.19)和(6.20),及(6.2),得

$$A^+ = A^+AA^+ = \left(A^+F\right)\left(R^*A^+\right) = R\left(R^*R\right)^{-1}\left(F^*F\right)^{-1}F^*.$$

特别,对于 A 满秩的情形:当 $r = m \le n$ 时,在(6.16)中取

$$F = I_m, R^* = A.$$

当 $r = n \le m$ 时,在(6.16)中取

$$F = A,\ \ R^* = I_n,$$

直接得(6.17). □

在具体计算时,对于秩分解(6.15),可以选取 A 的线性无关的 r 个列作为 F 的列;而 R^* 中相应的列选取适当的单位向量,然后再确定其余的列.

7.6.8 定理 设 A^+ 是 $A \in \mathbb{C}^{m\times n}$ 的 Moore-Penrose 逆,则

(1) 当 $A = 0$ 时, $A^+ = 0^{\mathrm{T}}$.

(2) $\left(A^+\right)^+ = A$.

(3) $(\alpha A)^+ = \alpha^{-1}A^+, \forall \alpha \in \mathbb{C}, \alpha \ne 0$.

(4) $\left(A^*\right)^+ = \left(A^+\right)^*$.

(5) $\left(AA^*\right)^+ = \left(A^+\right)^*A^+,\ \left(A^*A\right)^+ = A^+\left(A^+\right)^*$.

(6) 如果 $m = n$ 且 A 是正规矩阵,则 A^+ 也是正规矩阵,而且

$$\left(A^k\right)^+ = \left(A^+\right)^k,\quad k = 1,2,\cdots$$

(7) 对于任何酉矩阵 $U \in \mathbb{C}^{m\times m}, V \in \mathbb{C}^{n\times n}$,

$$\left(UAV\right)^+ = V^*A^+U^*.$$

(8) $A^+Ax = x, \forall x \in \mathscr{R}_{A^*}$; $AA^+y = y, \forall y \in \mathscr{R}_A$.

证 表示式(6.16)可以用来证明以上性质.

这里仅以性质(7)为例,示范证明的方法,其余性质的证明留作练习.

若 $A=FR^*$ 是 A 的秩分解,则 $(UF)(R^*V)$ 是 UAV 的秩分解.因此,利用(6.16)及 U 和 V 的酉性质,有

$$\begin{aligned}(UAV)^+ &= (V^*R)(R^*VV^*R)^{-1}(F^*UU^*F)^{-1}(F^*U^*)\\ &= V^*R(R^*R)^{-1}(F^*F)^{-1}F^*U^* = V^*A^+U^*.\end{aligned}$$ □

7.6.9 定理 设 $A\in\mathbb{C}^{m\times n}$,并设

$$\sigma_1\ge\cdots\ge\sigma_r>0,\quad \sigma_{r+1}=\cdots=\sigma_n=0$$

是 A 的奇异值,则 $\sigma_1^{-1},\cdots,\sigma_r^{-1}$ 是 A^+ 的非零奇异值.

而且,如果 $\boldsymbol{u}^{(1)},\cdots,\boldsymbol{u}^{(m)}$ 和 $\boldsymbol{v}^{(1)},\cdots,\boldsymbol{v}^{(n)}$ 分别是 A 的单位左和右奇异向量,那么 $\boldsymbol{v}^{(1)},\cdots,\boldsymbol{v}^{(n)}$ 和 $\boldsymbol{u}^{(1)},\cdots,\boldsymbol{u}^{(m)}$ 分别是 A^+ 的单位左和右奇异向量.

证 依奇异值分解定理 **5.4.5**,存在酉矩阵 $U\in\mathbb{C}^{m\times m}$ 和 $V\in\mathbb{C}^{n\times n}$,使得

$$U^*AV=\Sigma\equiv\begin{bmatrix}\Sigma_r & 0_3\\ 0_2 & 0_1\end{bmatrix},\ \Sigma_r=\mathrm{diag}(\sigma_1,\cdots,\sigma_r),\tag{6.21}$$

U 的各列 $\boldsymbol{u}^{(1)},\cdots,\boldsymbol{u}^{(m)}$ 和 V 的各列 $\boldsymbol{v}^{(1)},\cdots,\boldsymbol{v}^{(n)}$ 分别是 A 的单位左和右奇异向量.

再依(6.21)和 **7.6.8** 的(7),有

$$V^*A^+U=(U^*AV)^+=\Sigma^+=\begin{bmatrix}\Sigma_r^+ & 0_2^{\mathrm{T}}\\ 0_3^{\mathrm{T}} & 0_1^{\mathrm{T}}\end{bmatrix},$$

$$\Sigma_r^+=\Sigma_r^{-1}=\mathrm{diag}(\sigma_1^{-1},\cdots,\sigma_r^{-1})\tag{6.22}$$

这表明 $\sigma_1^{-1},\cdots,\sigma_r^{-1}$ 是 A^+ 的非零奇异值,且 V 的各列 $\boldsymbol{v}^{(1)},\cdots,\boldsymbol{v}^{(n)}$ 和 U 的各列 $\boldsymbol{u}^{(1)},\cdots,\boldsymbol{u}^{(m)}$ 分别是 A^+ 的单位左和右奇异向量. □

7.6.10 定理 设 $A\in\mathbb{C}^{m\times n}$, $b\in\mathbb{C}^m$,则

$$x^{(0)}\equiv A^+b\tag{6.23}$$

满足

$$\left\|Ax^{(0)}-b\right\|_2=\min_x\left\|Ax-b\right\|_2, \tag{6.24}$$

其中$\|\bullet\|_2$是$\mathbb{C}^n$中的 Euclid 范数. $x^{(0)}$称为方程组

$$Ax=b \tag{6.25}$$

在(6.24)意义下的**最佳逼近解**或称为**最小二乘解**.

证 对于任意的$x\in\mathbb{C}^n$,

$$Ax-b=A\left(x-A^+b\right)+\left(I-AA^+\right)(-b), \tag{6.26}$$

显然,

$$A\left(x-A^+b\right)\in\mathscr{R}_A.$$

另一方面,对于任何$y\in\mathscr{R}_A$,存在$z\in\mathbb{C}^n$,使得$y=Az$,于是依(6.3)和(6.1),有

$$\begin{aligned}\left(\left(I-AA^+\right)(-b),y\right)&=-(b,Az)+\left(AA^+b,Az\right)\\&=-(b,Az)+\left(b,\left(AA^+\right)^*Az\right)\\&=-(b,Az)+\left(b,AA^+Az\right)\\&=-(b,Az)+(b,Az)=0,\end{aligned}$$

推出

$$\left(I-AA^+\right)(-b)\in\mathscr{R}_A^\perp.$$

因此,可对(6.26)应用 Pythagoras 定理(见 **1.7.7**),得

$$\begin{aligned}\|Ax-b\|_2^2&=\left\|A\left(x-A^+b\right)\right\|_2^2+\left\|\left(I-AA^+\right)(-b)\right\|_2^2\\&=\left\|A\left(x-x^{(0)}\right)\right\|_2^2+\left\|Ax^{(0)}-b\right\|_2^2.\end{aligned}$$

这一关系式表明,当

$$x-x^{(0)}\in\mathscr{N}_A$$

时,特别,包括了取$x=x^{(0)}$,使$\|Ax-b\|_2$达到最小值

$$\left\|Ax^{(0)}-b\right\|_2. \qquad \square$$

这个定理给出了 Moore-Penrose 逆的一个很漂亮的应用.在处理不相容的或超定的方程组$Ax=b$时,最小二乘逼近是一个非常

有用的方法.

下一定理进一步提供了最小二乘解 $x^{(0)}$ 通过 A 的奇异值和奇异向量的显式表示.

7.6.11 定理 设 $A \in \mathbb{C}^{m\times n}, b \in \mathbb{C}^m$,并设

$$\sigma_1 \geq \cdots \geq \sigma_r > 0, \quad \sigma_{r+1} = \cdots = \sigma_n = 0$$

是 A 的奇异值,$v^{(1)}, \cdots, v^{(n)} \in \mathbb{C}^n$ 是相应的单位右奇异向量,构成 $\mathbb{C}^n$ 的一组基,则方程(6.25)的最小二乘解

$$x^{(0)} = \alpha_1 v^{(1)} + \cdots + \alpha_r v^{(r)}, \tag{6.27}$$

其中

$$\alpha_i = \frac{\left(A^* b, v^{(i)}\right)}{\sigma_i^2}, \quad i = 1, \cdots, r. \tag{6.28}$$

证 设 $u^{(1)}, \cdots, u^{(m)}$ 是 A 的单位左奇异向量,构成 $\mathbb{C}^m$ 的一组基,则 b 有表示式

$$b = \beta_1 u^{(1)} + \cdots + \beta_m u^{(m)},$$

其中

$$\beta_i = \left(b, u^{(i)}\right), \quad i = 1, \cdots, m.$$

于是,根据(6.23),有

$$x^{(0)} \equiv A^+ b = \sum_{i=1}^{m} \beta_i A^+ u^{(i)} = \sum_{i=1}^{r} \beta_i \sigma_i^{-1} v^{(i)}.$$

另一方面,从(6.21)看,

$$u^{(i)} = \frac{1}{\sigma_i} A v^{(i)}, \quad i = 1, \cdots, r,$$

因此

$$\alpha_i \equiv \beta_i \sigma_i^{-1} = \frac{1}{\sigma_i}\left(b, u^{(i)}\right) = \frac{1}{\sigma_i^2}\left(b, A v^{(i)}\right)$$

$$= \frac{1}{\sigma_i^2}\left(A^* b, v^{(i)}\right), \quad i = 1,\cdots,r. \qquad \square$$

7.6.12 定义 $A \in \mathbb{C}^{n\times n}$ 称为 **EP-矩阵**,如果满足

$$A^+ A = AA^+. \tag{6.29}$$

注意,依 **7.6.8** 的(1), A 为零矩阵时, A^+ 也是零矩阵;特别,视 $a \in \mathbb{C}$为 1×1 矩阵时,有

$$a^+ = \begin{cases} 1/a, & a \neq 0, \\ 0, & a = 0. \end{cases} \tag{6.30}$$

容易证明(留作练习):若 $A = \mathrm{diag}(\lambda_1,\cdots,\lambda_n)$,则

$$A^+ = \mathrm{diag}(\lambda_1^+,\cdots,\lambda_n^+).$$

7.6.13 定理 设 $A \in \mathbb{C}^{n\times n}$,下列条件等价:

(1) A 是 EP-矩阵.

(2) $Au = \lambda u$ 当且仅当 $A^+ u = \lambda^+ u$.

(3) $AX = 0$ 当且仅当 $A^* X = 0, X \in \mathbb{C}^{n\times n}$.

证明留作练习.

从(1)和(3)等价推出:任何自伴矩阵是 EP-矩阵.

7.7 (i,j,k)型逆

7.7.1 定义 设 $A \in \mathbb{C}^{m\times n}$,四个按顺序编号为(1),(2),(3),(4)的方程如下:

$$AXA = A, \qquad (1) \tag{7.1}$$

$$XAX = X, \quad (2) \qquad (7.2)$$

$$(AX)^* = AX, \quad (3) \qquad (7.3)$$

$$(XA)^* = XA. \quad (4) \qquad (7.4)$$

满足其某几个方程的矩阵 $X \in \mathbb{C}^{n\times m}$ 可称之为 A 的一种广义逆,如此规定的多种多样的广义逆统称 (i,j,k) **-逆**.

(1,2,3,4)**-逆**:指满足方程(7.1)~(7.4)的 X,亦即 Moore-Penrose 逆, $X = A^+$.

(i,j,k) **-逆**($1 \le i < j < k \le 4$):指满足(7.1)~(7.4)的第 i,第 j 和第 k 个方程的 X.这样的矩阵 X 并不唯一,记其全体组成的集合为

$$A\{i,j,k\},$$

并将 $A\{i,j,k\}$ 的每个成员称为 A 的 (i,j,k) **-逆**.特别,

(1,2,3)-逆也称为**最小平方自反 g-逆**或**正规 g-逆**;

(1,2,4)-逆也称为**最小范数自反 g-逆**(minimum-norm reflexive g-inverse)或**弱 g-逆**.

(i,j) **-逆**($1 \le i < j \le 4$):指满足(7.1)~(7.4)的第 i 和第 j 个方程的 X,记其全体组成的集合为

$$A\{i,j\},$$

并将 $A\{i,j\}$ 的每个成员称为 A 的 (i,j) **-逆**.特别,

(1,2)-逆也称为**自反 g-逆**;

$A\{1,2\}$ 的每个成员就是 A 的广义逆 A^{I};

(1,3)-逆也称为**最小平方 g-逆**;

(1,4)-逆也称为**最小范数 g-逆**.

(i) **-逆**($1 \le i \le 4$):指满足(7.1)~(7.4)的第 i 个方程的 X,记其全体组成的集合为

$$A\{i\},$$

并将 $A\{i\}$ 的每个成员称为 A 的 (i) **-逆**.特别,

(1)-逆也称为 **g-逆**或**伪逆**(pseudo inverse),记作 A^-.

"自反"一词是鉴于方程(7.1)和(7.2)的对称性:如果 X 是 A 的(1,2)-逆,那么 A 是 X 的(1,2)-逆.

注意,各种 (i,j,k) -逆大多尚没有一致认可的名称,包括上面所列的名称(见[RM1971]).

(1)-逆是一个广泛使用的概念.如同上述列举的术语所显示的那样,在 (i,j,k) -逆中有应用意义的各种逆多是 $A\{1\}$ 的成员.

7.7.2 定理 设 $A\in\mathbb{C}^{m\times n}$, $A^-\in\mathbb{C}^{n\times m}$ 是 A 的一个(1)-逆,则

$$A\{1\}=\{A^-+Y-A^-AYAA^-:Y\in\mathbb{C}^{n\times m}\}. \tag{7.5}$$

如果 $\operatorname{rank}A=r$,而且存在非奇异矩阵 $R\in\mathbb{C}^{m\times m}$, $Q\in\mathbb{C}^{n\times n}$,使得

$$A=R\begin{bmatrix} I_r & 0\\ 0 & 0\end{bmatrix}Q, \tag{7.6}$$

那么有另一种表示:

$$A\{1\}=$$

$$\left\{Q^{-1}\begin{bmatrix} I_r & U\\ V & W\end{bmatrix}R^{-1}:U\in\mathbb{C}^{r\times(m-r)},V\in\mathbb{C}^{(n-r)\times r},W\in\mathbb{C}^{(n-r)\times(m-r)}\right\}. \tag{7.7}$$

证 先证(7.5). A^- 是(1)-逆,满足

$$AA^-A=A. \tag{7.8}$$

令

$$A_Y=A^-+Y-A^-AYAA^-. \tag{7.9}$$

一方面,对任意的$Y \in \mathbb{C}^{n\times m}$,

$$AA_Y A = AA^- A + AYA - AA^- AYAA^- A$$
$$= A + AYA - AYA = A,$$

故$A_Y \in A\{1\}$.

另一方面,对任意的$B \in A\{1\}$,取

$$Y = B - A^-,$$

注意到$ABA = A$,有

$$A_{B-A^-} = A^- + B - A^- - A^- ABAA^- + A^- AA^- AA^-$$
$$= B - A^- AA^- + A^- AA^- = B,$$

表明B可以表示成(7.9)的形式.

现证(7.7).将(7.6)代入(7.8),并利用R和Q是可逆的,得

$$\begin{bmatrix} I_r & 0 \\ 0 & 0 \end{bmatrix} QA^- R \begin{bmatrix} I_r & 0 \\ 0 & 0 \end{bmatrix} = \begin{bmatrix} I_r & 0 \\ 0 & 0 \end{bmatrix}, \tag{7.10}$$

把$QA^- R$写成匹配的分块形式:

$$QA^- R = \begin{bmatrix} Z & U \\ V & W \end{bmatrix}, \quad Z \in \mathbb{C}^{r\times r}. \tag{7.11}$$

因为

$$\begin{bmatrix} I_r & 0 \\ 0 & 0 \end{bmatrix}\begin{bmatrix} Z & U \\ V & W \end{bmatrix}\begin{bmatrix} I_r & 0 \\ 0 & 0 \end{bmatrix} = \begin{bmatrix} Z & 0 \\ 0 & 0 \end{bmatrix},$$

所以由(7.10)推出$Z = I_r$.于是,从(7.11)有

$$A^- = Q^{-1}\begin{bmatrix} I_r & U \\ V & W \end{bmatrix} R^{-1}, \tag{7.12}$$

故(7.7)成立. □

7.7.3 定理 设$A \in \mathbb{C}^{m\times n}$, $A^- \in A\{1\}$,则

(1) $\left(A^-\right)^{\mathrm{T}} = \left(A^{\mathrm{T}}\right)^-$.

(2) 如果A是自伴矩阵,那么$\left(A^-\right)^* \in A\{1\}$.

(3) $A^{-}AA^{-} \in A\{1,2\}$.

证明留作练习.

7.7.4 定理 设 $A \in \mathbb{C}^{m\times n}$, $A^{-} \in A\{1\}$ 且表示成(7.12)的形式,则

(1) $A^{-} \in A\{1,2\}$的充分必要条件是

$$W = VU.$$

(2) $A^{-} \in A\{1,2,3\}$的充分必要条件是

$$U = 0, \quad W = 0.$$

(3) A^{-}为 Moore-Penrose 逆的充分必要条件是

$$U = 0,\ V = 0,\ W = 0.$$

证明留作练习.

7.7.5 定理 设 $A \in \mathbb{C}^{m\times n}$ 具有分块形式

$$A = \begin{bmatrix} A_1 & A_2 \\ A_3 & A_4 \end{bmatrix}, \tag{7.13}$$

其中 $A_1 \in \mathbb{C}^{r\times r}$ 是非奇异矩阵, $r = \mathrm{rank}A$; $A_4 \in \mathbb{C}^{(n-r)\times(m-r)}$,而且满足条件

$$A_4 = A_3 A_1^{-1} A_2, \tag{7.14}$$

则

$$\begin{bmatrix} A_1^{-1} & -A_1^{-1}A_2W \\ 0 & W \end{bmatrix} \in A\{1\}, \quad \forall W \in \mathbb{C}^{(n-r)\times(m-r)}. \tag{7.15}$$

证 考虑方程

$$\begin{bmatrix} A_1 & A_2 \\ A_3 & A_4 \end{bmatrix}\begin{bmatrix} X_1 & X_2 \\ X_3 & X_4 \end{bmatrix}\begin{bmatrix} A_1 & A_2 \\ A_3 & A_4 \end{bmatrix} = \begin{bmatrix} A_1 & A_2 \\ A_3 & A_4 \end{bmatrix}$$

的部分解,其中 $X_1 \in \mathbb{C}^{r\times r}$, $X_4 \in \mathbb{C}^{(n-r)\times(m-r)}$.为了容易求解,取

$$X_1 = A_1^{-1}, \qquad X_3 = 0,$$

并记 X_4 为 W,便可推出在条件(7.14)下,

$$X_2 = -A_1^{-1}A_2W. \qquad \square$$

此定理可以引出一些有意思的特性.

如果 A 是方阵,自然 W 也是方阵,取 W 是非奇异的,那么形如(7.15)的矩阵也是非奇异的.这说明,与 Moore-Penrose 逆不同,即便 A 是奇异的,其(1)-逆也有可能是非奇异的.

类似地,当 A 是自伴矩阵时,其(1)-逆或(1,2)-逆不必定是自伴的,仍然有别于 Moore-Penrose 逆的情形.

例 考虑矩阵

$$A = \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix}, \tag{7.16}$$

容易检验 A 是其自身的一个(1)-逆.利用(7.5),取 $A^- = A$,可得 A 的(1)-逆的一般表示

$$\begin{bmatrix} 1 & y_1 \\ y_2 & y_3 \end{bmatrix}, \quad \forall y_1, y_2, y_3 \in \mathbb{C}. \tag{7.17}$$

尽管(7.16)中的 A 是奇异的,只要 $y_3 \neq y_1 y_2$,(7.17)中的矩阵便是非奇异的.类似地,虽然(7.16)中的 A 是对称的,(7.17)中的矩阵却仅当 $y_1 = y_2$ 时才是对称的.

7.7.6 定义 n-终点网络(n-terminal network)的电流向量

$$i = (i_1, \cdots, i_n)^{\mathrm{T}}$$

和电压向量

$$v = (v_1, \cdots, v_n)^{\mathrm{T}}$$

的关系是

$$i = Yv,$$

其中 $Y \in \mathbb{R}^{n\times n}$ 是不定容许矩阵,满足

$$Ye = 0, \quad e^{\mathrm{T}}Y = 0, \quad e \equiv (1,\cdots,1)^{\mathrm{T}} \in \mathbb{R}^{n}, \tag{7.18}$$

Y 称为**双心矩阵**(doubly centred matrix).

显然,双心矩阵是奇异的.

7.7.7 定理 设 $Y \in \mathbb{R}^{n\times n}$ 是双心矩阵, $Z \equiv Y + \frac{1}{n}ee^{\mathrm{T}}$ 是非奇异矩阵,则

$$Z^{-1}Y = I - \frac{1}{n}ee^{\mathrm{T}}, \tag{7.19}$$

从而 $Z^{-1} \in Y\{1\}$.

证 依假设, Z 可逆.用 Z^{-1} 左乘 $Z \equiv Y + \frac{1}{n}ee^{\mathrm{T}}$,得

$$Z^{-1}Y = I - \frac{1}{n}Z^{-1}ee^{\mathrm{T}}. \tag{7.20}$$

另一方面,用 ee^{T} 右乘 $Z \equiv Y + \frac{1}{n}ee^{\mathrm{T}}$,注意到(7.18),有

$$Zee^{\mathrm{T}} = Yee^{\mathrm{T}} + \frac{1}{n}ee^{\mathrm{T}}ee^{\mathrm{T}} = \frac{1}{n}e\left(e^{\mathrm{T}}e\right)e^{\mathrm{T}} = ee^{\mathrm{T}}$$

等价于

$$Z^{-1}ee^{\mathrm{T}} = ee^{\mathrm{T}}, \tag{7.21}$$

将(7.21)代入(7.20)即得(7.19).

最后,由(7.19)及(7.18)即可推出 $YZ^{-1}Y = Y$,从而 $Z^{-1} \in Y\{1\}$. □

7.7.8 定理 设 $Y \in \mathbb{R}^{n\times n}$ 是双心矩阵, $Z \equiv Y + \frac{1}{n}ee^{\mathrm{T}}$ 是非奇异矩阵, $X \equiv Z^{-1} - \frac{1}{n}ee^{\mathrm{T}}$,则

$$XY = YX = I - \frac{1}{n}ee^{\mathrm{T}}, \tag{7.22}$$

从而 $X = Y^{+}$,而且 X 也是双心矩阵.

证 应用(7.18)和(7.19),有

$$XY = Z^{-1}Y - \frac{1}{n}ee^{\mathrm{T}}Y = Z^{-1}Y = I - \frac{1}{n}ee^{\mathrm{T}}.$$

又仿关于(7.19)的推证,可得 $YZ^{-1} = I - \frac{1}{n}ee^{\mathrm{T}}$,于是

$$YX = YZ^{-1} - \frac{1}{n}Yee^{\mathrm{T}} = YZ^{-1} = I - \frac{1}{n}ee^{\mathrm{T}},$$

因此(7.22)成立.而且

$$YXY = Y, \quad (XY)^{*} = XY, \quad (YX)^{*} = YX.$$

又注意到(7.21),得

$$XYX = X - Xee^{\mathrm{T}} = X - \left(Z^{-1} - \frac{1}{n}ee^{\mathrm{T}}\right)ee^{\mathrm{T}}$$

$$= X - \left(Z^{-1}ee^{\mathrm{T}} - ee^{\mathrm{T}}\right) = X,$$

表明 $X = Y^{+}$.最后,因

$$Ze = Ye + \frac{1}{n}e\left(e^{\mathrm{T}}e\right) = e,$$

等价于 $Z^{-1}e = e$,故

$$Xe = Z^{-1}e - \frac{1}{n}e\left(e^{\mathrm{T}}e\right) = Z^{-1}e - e = 0$$

同样,有 $e^{\mathrm{T}}X = 0$.所以 X 也是双心矩阵. □

7.8 Drazin 逆

7.8.1 定义 设 $A \in \mathbb{C}^{n\times n}$.矩阵 $A^{\mathrm{D}} \in \mathbb{C}^{n\times n}$ 称为 A 的 **Drazin 逆**或 **D-**

逆,如果成立

$$A^{k+1}A^{\mathrm{D}} = A^{k}, \tag{8.1}$$

$$A^{\mathrm{D}}AA^{\mathrm{D}} = A^{\mathrm{D}}, \tag{8.2}$$

$$AA^{\mathrm{D}} = A^{\mathrm{D}}A, \tag{8.3}$$

其中k是A的指数,是使得$\mathrm{rank}A^{k} = \mathrm{rank}A^{k+1}$的最小非负整数.注意,若$A$非奇异,则$k=0$,并约定$A^{0}=I$,从而$A^{\mathrm{D}} = A^{-1}$.

(8.3)是说,如同通常的逆那样,A^{D}和A可以交换,而 Moore-Penrose 逆和其它广义逆一般不具有如此性质.

但是,一般地,A^{D}并不满足(7.1)中的方程$AXA=A$,故$A^{\mathrm{D}} \notin A\{1\}$.因而 D-逆和$(i,j,k)$-逆有较大的区别,它应用于有限链及其它一些方面.

7.8.2 引理 设$A \in \mathbb{C}^{n\times n}$,$A^{\mathrm{D}}$是$A$的 D-逆,则

(1) $A^{\mathrm{D}}A^{k+1} = A^{k}$.

(2) $\left(A^{\mathrm{D}}\right)^{l+1}A^{l} = A^{\mathrm{D}} = A^{l}\left(A^{\mathrm{D}}\right)^{l+1}, l=0,1,\cdots$

证 利用(8.1)和(8.3),得

$$\begin{aligned} A^{k} &= A^{k+1}A^{\mathrm{D}} = A^{k}\left(AA^{\mathrm{D}}\right) = A^{k}A^{\mathrm{D}}A = A^{k-1}\left(AA^{\mathrm{D}}\right)A \\ &= A^{k-1}A^{\mathrm{D}}A^{2} = \cdots = A^{\mathrm{D}}A^{k+1}. \end{aligned}$$

利用(8.2)和(8.3),有

$$A^{\mathrm{D}} = A^{\mathrm{D}}\left(AA^{\mathrm{D}}\right) = \left(A^{\mathrm{D}}\right)^{2}A.$$

然后重复使用此关系式,

$$\begin{aligned} A^{\mathrm{D}} &= A^{\mathrm{D}}\left((A^{\mathrm{D}})^{2}A\right)A = \left(A^{\mathrm{D}}\right)^{2}\left((A^{\mathrm{D}})^{2}A\right)A^{2} \\ &= \left(A^{\mathrm{D}}\right)^{3}\left((A^{\mathrm{D}})^{2}A\right)A^{3} = \cdots = \left(A^{\mathrm{D}}\right)^{l+1}A^{l}, \ l=0,1,\cdots \end{aligned}$$

类似地,使用关系式

$$A^{\mathrm{D}} = \left(AA^{\mathrm{D}}\right)A^{\mathrm{D}} = A\left(A^{\mathrm{D}}\right)^{2},$$

即得

$$A^{\mathrm{D}} = A^{l}\left(A^{\mathrm{D}}\right)^{l+1}, \ l=0,1,\cdots$$

□

7.8.3 定理 对于任何 $A \in \mathbb{C}^{n\times n}$ 存在唯一的 D-逆 $A^{\mathrm{D}} \in \mathbb{C}^{n\times n}$.

证 依 **5.1.2**, A 相似于其 Jordan 标准形 J,可以表示成

$$A = PJP^{-1}, \quad J = \begin{bmatrix} C & 0 \\ 0 & N \end{bmatrix}, \tag{8.4}$$

其中 $P \in \mathbb{C}^{n\times n}$ 和 $C \in \mathbb{C}^{r\times r}$ 是非奇异矩阵, $N \in \mathbb{C}^{(n-r)\times(n-r)}$ 是指数为 k 的幂零矩阵, $k \le n$ 是某个非负整数.注意到

$$N^{k-1} \ne 0, \quad N^k = N^{k+1} = \cdots = 0,$$

方程(8.1)成为

$$P\begin{bmatrix} C^{k+1} & 0 \\ 0 & 0 \end{bmatrix}P^{-1}A^{\mathrm{D}} = A^{k+1}A^{\mathrm{D}} = A^k = P\begin{bmatrix} C^k & 0 \\ 0 & 0 \end{bmatrix}P^{-1},$$

由此,有

$$\begin{bmatrix} C^{k+1} & 0 \\ 0 & 0 \end{bmatrix}P^{-1}A^{\mathrm{D}}P = \begin{bmatrix} C^k & 0 \\ 0 & 0 \end{bmatrix}. \tag{8.5}$$

将 $P^{-1}A^{\mathrm{D}}P$ 写成匹配的分块形式:

$$P^{-1}A^{\mathrm{D}}P = \begin{bmatrix} Z & U \\ V & W \end{bmatrix}, \tag{8.6}$$

其中 $Z \in \mathbb{C}^{r\times r}, W \in \mathbb{C}^{(n-r)\times(n-r)}$;而且在等式(8.5)两端左乘矩阵 $\mathrm{diag}\left(\left(C^{-1}\right)^{k+1}, I_{n-r}\right)$,得

$$\begin{bmatrix} I_r & 0 \\ 0 & 0 \end{bmatrix}\begin{bmatrix} Z & U \\ V & W \end{bmatrix} = \begin{bmatrix} C^{-1} & 0 \\ 0 & 0 \end{bmatrix},$$

由此推出

$$Z = C^{-1}, \quad U = 0.$$

类似地,利用 **7.8.2** 的(1),有

$$P^{-1}A^{\mathrm{D}}P\begin{bmatrix} C^{k+1} & 0 \\ 0 & 0 \end{bmatrix} = \begin{bmatrix} C^k & 0 \\ 0 & 0 \end{bmatrix},$$

从而又可推出

$$V=0,$$

于是(8.6)成为

$$P^{-1}A^{\mathrm{D}}P=\begin{bmatrix}C^{-1}&0\\0&W\end{bmatrix}\quad\text{或}\quad A^{\mathrm{D}}=P\begin{bmatrix}C^{-1}&0\\0&W\end{bmatrix}P^{-1}.$$

再利用 **7.8.2** 的(2)及(8.4),

$$\begin{aligned}A^{\mathrm{D}}&=\left(A^{\mathrm{D}}\right)^{k+1}A^{k}\\&=P\begin{bmatrix}\left(C^{-1}\right)^{k+1}&0\\0&W^{k+1}\end{bmatrix}P^{-1}P\begin{bmatrix}C^{k}&0\\0&N^{k}\end{bmatrix}P^{-1},\end{aligned}$$

注意到$N^k=0$,有

$$\begin{aligned}\begin{bmatrix}C^{-1}&0\\0&W\end{bmatrix}&=P^{-1}A^{\mathrm{D}}P\\&=\begin{bmatrix}\left(C^{-1}\right)^{k+1}&0\\0&W^{k+1}\end{bmatrix}\begin{bmatrix}C^{k}&0\\0&N^{k}\end{bmatrix}=\begin{bmatrix}C^{-1}&0\\0&0\end{bmatrix},\end{aligned}$$

因此,$W=0$,得A的 D-逆为

$$A^{\mathrm{D}}=P\begin{bmatrix}C^{-1}&0\\0&0\end{bmatrix}P^{-1}.\tag{8.7}$$

现在证明A^{D}的唯一性.设$\hat{A}^{\mathrm{D}}$也满足(8.1)~(8.3),则由

$$A^{k+1}A^{\mathrm{D}}=A^{k},\quad A^{k+1}\hat{A}^{\mathrm{D}}=A^{k},$$

有

$$A^{k+1}\left(\hat{A}^{\mathrm{D}}-A^{\mathrm{D}}\right)=0.$$

由此,仿照上面的证明,便可推出

$$\hat{A}^{\mathrm{D}}-A^{\mathrm{D}}=0.$$

□

7.8.4 推论 $A^{\mathrm{D}}\in A\{1\}$的充分必要条件是(**7.8.1** 中的)$k\le 1$.

证 $A^{\mathrm{D}}\in A\{1\}$,依 **7.7.1**,满足$AA^{\mathrm{D}}A=A$.利用(8.4)和(8.7),即

成立

$$P\begin{bmatrix} C & 0 \\ 0 & N \end{bmatrix}P^{-1}P\begin{bmatrix} C^{-1} & 0 \\ 0 & 0 \end{bmatrix}P^{-1}P\begin{bmatrix} C & 0 \\ 0 & N \end{bmatrix}P^{-1} = P\begin{bmatrix} C & 0 \\ 0 & N \end{bmatrix}P^{-1},$$

简化后得 $N = 0$.

因此,N 或为指数 $k = 0$ 的幂零矩阵,或为指数 $k = 1$ 的幂零矩阵.

N 为 $k = 0$ 的幂零矩阵,即 N 是零阶矩阵,此时

$$A = PCP^{-1}$$

是非奇异矩阵,$A^{\mathrm{D}} = A^{-1}$.

N 为 $k = 1$ 的幂零矩阵, 即 N 是零矩阵,此时

$$A = P\mathrm{diag}(C,0)P^{-1},$$

A^{D} 必形如(8.7).

因此,$A^{\mathrm{D}} \in A\{1\}$等价于 $k \le 1$. □

此推论表明,$k = 0$ 或 1 时的 D-逆是(1,2)-逆,是(1,2)-逆中唯一可以和 A 交换的特殊情形.对这种情形使用术语**群逆**(group inverse),原因是 A 属于具有乘法逆的乘法群,并采用记号 $A^{\#}$.在这种情形下,A 可以表示成(8.4)的形式,其中 $N = 0$;$A^{\#}$ 仍同(8.7)所示.

$A^{\#}$ 在统计学中有着应用.

7.8.5 定理 设 $A \in \mathbb{C}^{n\times n}$,$A^{\mathrm{D}} = A^{+}$ 的充分必要条件是 A 为 EP-矩阵.

证明留作练习.

7.8.6 定理 设 $A \in \mathbb{C}^{n\times n}$,则 A 和 A^{D} 两者的非零特征值互为倒数.

证 设 A 有特征值 $\lambda \ne 0$ 及相应的特征向量 u;将 A 表示成(8.4)的形式,把 u 写成

$$u = P\begin{bmatrix} u^{(1)} \\ u^{(2)} \end{bmatrix}, \tag{8.8}$$

并把它们代入 $Au = \lambda u$,得

$$P\begin{bmatrix} Cu^{(1)} \\ Nu^{(2)} \end{bmatrix} = \lambda P\begin{bmatrix} u^{(1)} \\ u^{(2)} \end{bmatrix}.$$

鉴于 P 是非奇异矩阵,简化为

$$Cu^{(1)} = \lambda u^{(1)}, \quad Nu^{(2)} = \lambda u^{(2)}.$$

因为 N 的所有特征值为零及 C 是非奇异矩阵,推出 $u^{(2)} = 0$,而且

$$C^{-1}u^{(1)} = \lambda^{-1}u^{(1)}.$$

由此并利用(8.7),成立

$$A^{D}u = P\begin{bmatrix} C^{-1} & 0 \\ 0 & 0 \end{bmatrix}\begin{bmatrix} u^{(1)} \\ 0 \end{bmatrix} = \frac{1}{\lambda}\begin{bmatrix} u^{(1)} \\ 0 \end{bmatrix},$$

表明 A^{D} 以 λ^{-1} 和 u 作为特征值和相应的特征向量.如此性质,对广义特征向量也成立. □

通常逆矩阵 A^{-1} 和 A 两者的特征值(必为非零)是互为倒数的.上面定理说明 A^{D} 的非零特征值也具有通常逆矩阵的这种性质.但是, A^{+} 的非零特征值一般不是 A 的特征值的倒数.

8 矩阵分裂和迭代矩阵

8.1 矩阵迭代的基本原理

8.1.1 引言 凡线性模型若需给予数值答案,终究是求解线性代数方程组.

设线性方程组为

$$Ax = b, \tag{1.1}$$

其中 $A \in \mathbb{C}^{n \times n}$ 是已知非奇异矩阵,$b \in \mathbb{C}^n$ 是已知向量,$x \in \mathbb{C}^n$ 是待求解的向量.本章提供求解(1.1)的基本迭代法中有关矩阵的结构和性质.

一般,一个算法,假若其所有算术运算能精确进行,则能在有限运算次数内得出(1.1)的精确解,便称之为**直接法**;一个算法,即使其所有算术运算能精确进行,产生的只是趋向于(1.1)的精确解的一个近似解序列,便称之为**迭代法**.

迭代法的一般表示为

$$x^{(k+1)} = \Phi_k\left(x^{(k-m+1)}, \cdots, x^{(k)}\right), \quad k = m-1, m, \cdots \tag{1.2}$$

其中

$$\Phi_k : \underbrace{\mathbb{C}^n \times \cdots \times \mathbb{C}^n}_{m\text{个}} \to \mathbb{C}^n, \quad k = m-1, m, \cdots$$

通常由矩阵 A 及向量 b 导出,称为**迭代算子**. $x^{(0)}, \cdots, x^{(m-1)}$ 称为**迭代初值**,可以任意给定或按某种估计选取.递推关系(1.2)依赖前 m 步信息,产生向量序列 $\{x^{(k)}\}$,称为 m **步迭代法**. 1 步迭代法称为**单步迭**

代法.如果

$$\Phi_k \equiv \Phi, \quad k = m-1, m, \cdots$$

即迭代算子与 k 无关,则(1.2)称为**定常迭代法**;否则称为**不定常迭代法**.

单步、定常而且线性的迭代法是基本的,它们的统一形式如下:

$$x^{(k+1)} = Gx^{(k)} + h, \quad k = 0,1,2,\cdots \tag{1.3}$$

其中 $G \in \mathbb{C}^{n\times n}$ 由 A 导出,称为相伴 A 的**迭代矩阵**, $x^{(0)}$ 称为**初值**.

下面主要讨论此类迭代法.

8.1.2 定义 如果存在 $x^* \in \mathbb{C}^n$,使得对于任意的初值 $x^{(0)} \in \mathbb{C}^n$,由迭代法(1.3)产生的序列 $\{x^{(k)}\}$ 都收敛到 x^*,即

$$\lim_{k\to\infty} x^{(k)} = x^*,$$

则称迭代法(1.3)是**收敛**的,并以 x^* 为**极限**;否则,称其是**发散**的.

显然,如果迭代法收敛到极限 $x^* \in \mathbb{C}^n$,那么成立

$$x^* = Gx^* + h. \tag{1.4}$$

因此,若存在非奇异矩阵 $Q \in \mathbb{C}^{n\times n}$,使得

$$I - G = QA, \quad h = Qb, \tag{1.5}$$

则(1.4)等价于 $Ax^* = b$,即 x^* 是方程组(1.1)的解.

这说明,一个有用的形如(1.3)的迭代法,最起码的要求是满足条件(1.5).

通常,当条件(1.5)成立时,称迭代法(1.3)和方程组(1.1)是**相容**的.

现在,要考虑的问题是:怎样构造相容的迭代法(主要是迭代矩阵的结构)怎样判断迭代法的收敛性;怎样评价迭代法收敛的快慢,即所谓收敛速度.

关于构造相容的迭代法,第一步是按适当结构将矩阵 A **分裂**成两部分:

$$A = M - N, \tag{1.6}$$

其中 M 是非奇异矩阵.

然后,将(1.6)代入(1.1),可建立迭代过程

$$Mx^{(k+1)} = Nx^{(k)} + b, \quad k = 0,1,2,\cdots \tag{1.7}$$

只要取

$$G = M^{-1}N, \ h = M^{-1}b,$$

便可以将(1.7)写成(1.3)的形式.

显然,如此迭代法必和(1.1)相容.

实际计算时,无论为了计算 $M^{-1}N$ 和 $M^{-1}b$ 以采用迭代形式(1.3),或者为了保证精度等原因,宁可直接利用(1.7),每迭代一步解一个以 M 为系数矩阵的线性方程组,都必须选取 M 具有特殊结构,诸如对角或块对角,三角或块三角等,以使其求逆或求解相应方程组易于实现.

对 A 的不同的分裂,引出不同的迭代法.

8.1.3 定理 下列三个条件等价:

(1) 迭代法(1.3)收敛.

(2) $\lim\limits_{k\to\infty} G^k = 0$.

(3) $\rho(G) < 1$.

证 (1)$\Leftrightarrow$(2):设 $x^* \in \mathbb{C}^n$ 是方程组 $x = Gx + h$ 的解,即成立(1.4).

将(1.3)和(1.4)相减,得

$$x^{(k+1)} - x^* = G\left(x^{(k)} - x^*\right), \quad k = 0,1,2,\cdots \tag{1.8}$$

由此递推,有

$$x^{(k+1)} - x^* = G\left(x^{(k)} - x^*\right) = G^2\left(x^{(k-1)} - x^*\right)$$
$$= \cdots = G^{k+1}\left(x^{(0)} - x^*\right),$$

或者写成

$$\left(x^{(k)} - x^*\right) = G^k\left(x^{(0)} - x^*\right), \quad k = 0,1,2,\cdots \tag{1.9}$$

这表明

$$\lim_{k\to\infty} x^{(k)} = x^* \Leftrightarrow \lim_{k\to\infty} G^k = 0, \quad \forall x^{(0)} \in \mathbb{C}^n.$$

(2) $\Leftrightarrow$ (3):见 **2.1.23**. □

8.1.4 定理 设 $\|G\| < 1$,则迭代法(1.3)产生的序列 $\{x^{(k)}\}$ 收敛,而且对于 $k = 1,2,\cdots$,成立

$$\left\|x^{(k)} - x^*\right\| \le \frac{\|G\|}{1 - \|G\|}\left\|x^{(k)} - x^{(k-1)}\right\| \tag{1.10}$$

和

$$\left\|x^{(k)} - x^*\right\| \le \frac{\|G\|^k}{1 - \|G\|}\left\|x^{(1)} - x^{(0)}\right\|, \tag{1.11}$$

其中 $x^* = \lim\limits_{k\to\infty} x^{(k)}$,矩阵范数是向量范数导出的算子范数.

证 利用(1.8),

$$\begin{aligned}\left\|x^{(k)} - x^*\right\| &\le \|G\|\left\|x^{(k-1)} - x^*\right\| \\ &= \|G\|\left\|x^{(k-1)} - x^{(k)} + x^{(k)} - x^*\right\| \\ &\le \|G\|\left\|x^{(k)} - x^{(k-1)}\right\| + \|G\|\left\|x^{(k)} - x^*\right\|,\end{aligned}$$

由此并注意到 $\|G\| < 1$,推出(1.10).

类似地,利用(1.9),

$$\left\|x^{(k)} - x^*\right\| \le \left\|G^k\right\|\left\|x^{(0)} - x^*\right\| \le \|G\|^k\left\|x^{(0)} - x^*\right\|, \tag{1.12}$$

再利用(1.8),

$$\left\|x^{(0)} - x^*\right\| = \left\|x^{(0)} - x^{(1)} + x^{(1)} - x^*\right\|$$

$$\leq \left\|x^{(1)}-x^{(0)}\right\|+\left\|x^{(1)}-x^{*}\right\|$$
$$\leq \left\|x^{(1)}-x^{(0)}\right\|+\|G\|\left\|x^{(0)}-x^{*}\right\|,$$

推出

$$\left\|x^{(0)}-x^{*}\right\| \leq \frac{1}{1-\|G\|}\left\|x^{(1)}-x^{(0)}\right\|.$$

将此式代入(1.12)即得(1.11). □

8.1.5 定义 设$\|G\|<1$,称

$$R_k(G) \equiv -\ln\left(\left\|G^k\right\|^{\frac{1}{k}}\right)=-\frac{\ln\left\|G^k\right\|}{k} \tag{1.13}$$

为迭代法(1.3)的(*k* **次迭代的)平均收敛速度**;称

$$R(G) \equiv \lim_{k\to\infty} R_k(G)=-\ln \rho(G) \tag{1.14}$$

为迭代法(1.3)的**渐近收敛速度**.(1.14)中的后一等号见 **2.1.26**.

收敛速度可以有不同的定义,它们反映收敛的迭代法的收敛快慢,是一个重要的概念.

$R_k(G)$的意义如下:当$\|G\|<1$时,由迭代法(1.3)产生的向量序列$\left\{x^{(k)}\right\}$收敛到方程组$x=Gx+h$的解x^*.记

$$\varepsilon^{(k)}=x^{(k)}-x^{*}, \quad k=0,1,2,\cdots, \tag{1.15}$$

$\varepsilon^{(k)}$是第k次迭代所得$x^{(k)}$的误差向量.从(1.9),有

$$\left\|\varepsilon^{(k)}\right\| \leq\left\|G^k\right\|\left\|\varepsilon^{(0)}\right\|,$$

于是,得到$\left\|G^k\right\|$是k次迭代后$x^{(k)}$的误差和初值$x^{(0)}$的误差之比的一个上界,

$$\frac{\left\|\varepsilon^{(k)}\right\|}{\left\|\varepsilon^{(0)}\right\|} \leq\left\|G^{(k)}\right\|=\left(\left\|G^k\right\|^{\frac{1}{k}}\right)^k. \tag{1.16}$$

可以这样来解说该估计式:每次迭代使误差得以压缩,平均地看,其压缩率正比于$\|G^k\|^{1/k}$.显然,随迭代次数k的增加,$\|G^k\|$和$1/k$都是很小的正数.为了便于计算,自然地采用了(1.13)的定义方式.

但是,平均收敛速度$R_k(G)$依赖于迭代次数k及范数的选取,在理论分析和实际应用上常有不便,因此取其极限,$R(G)$是随迭代次数增加而渐近的误差的压缩率.

8.1.6 引理 设$A\in\mathbb{C}^{n\times n}$是自伴矩阵,且分裂为$A=M-N$,$M$为非奇异矩阵.则$M^*+N$是自伴矩阵,且对任意的$x\in\mathbb{C}^n$成立

$$x^*Ax-\tilde{x}^*A\tilde{x}=u^*\left(M^*+N\right)u, \tag{1.17}$$

其中$\tilde{x}=M^{-1}Nx,u=x-\tilde{x}$.

证 因为

$$\begin{aligned}M^*+N&=\left(A+N\right)^*+N=A+N^*+N\\&=M+N^*=\left(M^*+N\right)^*,\end{aligned}$$

所以M^*+N是自伴矩阵,而且

$$M=M^*-N^*+N. \tag{1.18}$$

注意到$M\tilde{x}=Nx$,有

$$Mu=Mx-M\tilde{x}=Mx-Nx=Ax$$

和

$$Nu=Nx-N\tilde{x}=M\tilde{x}-N\tilde{x}=A\tilde{x},$$

因此

$$x^*Ax-\tilde{x}^*A\tilde{x}=x^*Mu-\tilde{x}^*Nu,$$

由此并利用(1.18)及$x^*N^*=\tilde{x}^*M^*$,推出(1.17). □

8.1.7 定理 设$A\in\mathbb{C}^{n\times n}$是自伴矩阵,且分裂为$A=M-N$,$M$为非奇异矩阵,则

(1) 如果A和M^*+N正定,那么$\rho\left(M^{-1}N\right)<1$.

(2) 如果$\rho\left(M^{-1}N\right)<1$,$M^*+N$正定,那么$A$正定.

证 设$\lambda \in \lambda\left(M^{-1}N\right)$,则存在$x \in \mathbb{C}^n$, $x \neq 0$,使得

$$M^{-1}Nx = \lambda x. \tag{1.19}$$

令

$$\tilde{x} \equiv M^{-1}Nx = \lambda x, u \equiv x - \tilde{x} = (1-\lambda)x,$$

应用 **8.1.6**,有

$$\left(1-|\lambda|^2\right)x^* Ax = |1-\lambda|^2 x^*\left(M^* + N\right)x. \tag{1.20}$$

注意到若A正定,则因$x \neq 0$,有$Ax \neq 0$即$Mx \neq Nx$,表明$\lambda \neq 1$.于是,在A和$M^* + N$正定的假设下,从(1.20)知$|\lambda| < 1$;再由$\lambda \in \lambda\left(M^{-1}N\right)$的任意性,推出$\rho\left(M^{-1}N\right) < 1$.(1)得证.

现证(2).用反证法.

假若A非正定,则因$\rho\left(M^{-1}N\right) < 1$蕴涵$A$非奇异,必有$x^{(0)} \in \mathbb{C}^n$,使得

$$\eta \equiv \left(x^{(0)}\right)^* Ax^{(0)} < 0. \tag{1.21}$$

从$x^{(0)}$出发,构造序列

$$x^{(k+1)} = M^{-1}Nx^{(k)}, \quad k = 0,1,2,\cdots$$

因$\rho\left(M^{-1}N\right) < 1$,知

$$\lim_{k\to\infty} x^{(k)} = \lim_{k\to\infty}\left(M^{-1}N\right)^k x^{(0)} = 0. \tag{1.22}$$

令

$$u^{(k)} = x^{(k)} - x^{(k-1)}, k = 1,2,\cdots$$

应用 **8.1.6**,有

$$\left(x^{(k-1)}\right)^* Ax^{(k-1)} - \left(x^{(k)}\right)^* Ax^{(k)} = \left(u^{(k)}\right)^*\left(M^* + N\right)u^{(k)},$$
$$k = 1,2,\cdots \tag{1.23}$$

由于

$$Mu^{(1)} = -Ax^{(0)},$$

$x^{(0)}$满足(1.21),必有$u^{(1)} \neq 0$.因此,根据(1.23)及$M^* + N$正定,

$$\left(x^{(1)}\right)^* Ax^{(1)} = \left(x^{(0)}\right)^* Ax^{(0)} - \left(u^{(1)}\right)^*\left(M^* + N\right)u^{(1)}$$

$$< \left(x^{(0)}\right)^* A x^{(0)} = \eta < 0 .$$

相仿,可证对于 $k = 1,2,\cdots$,成立

$$\left(x^{(k)}\right)^* A x^{(k)} < \left(x^{(k-1)}\right)^* A x^{(k-1)} \le \eta < 0 ,$$

这与(1.22)矛盾. □

8.1.8 引理 设 $M = [m_{ij}] \in \mathbb{C}^{n\times n}$ 是严格对角优势矩阵,并设 $N = [n_{ij}] \in \mathbb{C}^{n\times n}$,则

$$\left\|M^{-1}N\right\|_\infty \le \max_{1\le i\le n} \frac{\sum_{j=1}^{n}\left|n_{ij}\right|}{\left|m_{ii}\right| - \sum_{j=1, j\ne i}^{n}\left|m_{ij}\right|}. \tag{1.24}$$

证 由矩阵算子范数定义知,必存在

$$y = \left(y_1,\cdots,y_n\right)^{\mathrm{T}} \in \mathbb{C}^n, \quad \|y\|_\infty = 1 ,$$

使得

$$\left\|M^{-1}N\right\|_\infty = \left\|M^{-1}Ny\right\|_\infty .$$

令

$$x \equiv \left(x_1,\cdots,x_n\right)^{\mathrm{T}} = M^{-1}Ny, \quad \left|x_{i_0}\right| \equiv \max_{1\le i\le n}\left|x_i\right| = \|x\|_\infty ,$$

则由 $Mx = Ny$,得

$$\sum_{j=1}^{n} m_{i_0 j} x_j = \sum_{j=1}^{n} n_{i_0 j} y_j .$$

于是,从 M 是严格对角优势矩阵,有

$$\left|x_{i_0}\right|\left(\left|m_{i_0 i_0}\right| - \sum_{j=1, j\ne i_0}^{n}\left|m_{i_0 j}\right|\right) \le \left(\left|m_{i_0 i_0} x_{i_0}\right| - \sum_{j=1, j\ne i_0}^{n}\left|m_{i_0 j} x_j\right|\right)$$

$$\le \left|\sum_{j=1}^{n} m_{i_0 j} x_j\right| = \left|\sum_{j=1}^{n} n_{i_0 j} y_j\right| \le \sum_{j=1}^{n}\left|n_{i_0 j} y_j\right| \le \sum_{j=1}^{n}\left|n_{i_0 j}\right| .$$

因此

$$\|M^{-1}N\|_\infty = \|x\|_\infty = |x_{i_0}|$$

$$\le \frac{\sum_{j=1}^{n}|n_{i_0 j}|}{|m_{i_0 i_0}| - \sum_{j=1, j\ne i_0}^{n}|m_{i_0 j}|} \le \max_{1\le i\le n} \frac{\sum_{j=1}^{n}|n_{ij}|}{|m_{ii}| - \sum_{j=1, j\ne i}^{n}|m_{ij}|}. \qquad \square$$

8.1.9 引理 设 $M=[m_{ij}]\in\mathbb{R}^{n\times n}$ 是严格对角优势的 L-矩阵,并设 $N=[n_{ij}]\in\mathbb{R}^{n\times n}$ 是非负矩阵,则

$$\rho(M^{-1}N) \ge \min_{1\le i\le n} \frac{\sum_{j=1}^{n}|n_{ij}|}{m_{ii} - \sum_{j=1, j\ne i}^{n}|m_{ij}|}. \qquad (1.25)$$

证 令

$$y \equiv (y_1,\cdots,y_n)^{\mathrm{T}} = M^{-1}Ne,$$

其中 $e=(1,\cdots,1)^T\in\mathbb{R}^n$.由定理假设,并依 **4.6.5**,知 $M^{-1}\ge 0$,从而

$$M^{-1}N\ge 0, \quad y\ge 0.$$

设 $y_{i_0}=\min\limits_{1\le i\le n} y_i$.从 $My=Ne$,并且注意到 L-矩阵的定义(见 **4.6.3**)有

$$\sum_{j=1}^{n} n_{i_0 j} = \sum_{j=1}^{n} m_{i_0 j} y_j = m_{i_0 i_0} y_{i_0} - \sum_{j=1, j\ne i_0}^{n} |m_{i_0 j}| y_j$$

$$\le y_{i_0}\left(m_{i_0 i_0} - \sum_{j=1, j\ne i_0}^{n} |m_{i_0 j}|\right). \qquad (1.26)$$

记 $M^{-1}N=[\beta_{ij}]$.依 **4.1.8**,并利用(1.26),得

$$\rho\left(M^{-1}N\right) \geq \min_{1\leq i\leq n}\sum_{j=1}^{n}\beta_{ij} = \min_{1\leq i\leq n} y_i = y_{i_0}$$

$$\geq \frac{\sum_{j=1}^{n} n_{i_0 j}}{m_{i_0 i_0} - \sum_{j=1, j\neq i_0}^{n}\left|m_{i_0 j}\right|} \geq \min_{1\leq i\leq n}\frac{\sum_{j=1}^{n}\left|n_{ij}\right|}{m_{ii} - \sum_{j=1, j\neq i}^{n}\left|m_{ij}\right|}. \qquad \square$$

8.1.10 定义 设线性方程组(1.1)中,系数矩阵 $A \in \mathbb{C}^{n\times n}$ 是非奇异的,已知向量 $b \in \mathbb{C}^n$,待求解向量 $x \in \mathbb{C}^n$.

在本章中统一使用如下记号:

$$D \equiv \operatorname{diag}\left(a_{11}, \cdots, a_{nn}\right), \tag{1.27}$$

$$C_L \equiv -\begin{bmatrix} 0 & \cdots & \cdots & 0 \\ a_{21} & \ddots & & \vdots \\ \vdots & \ddots & \ddots & \vdots \\ a_{n1} & \cdots & a_{n,n-1} & 0 \end{bmatrix}, \tag{1.28}$$

$$C_U \equiv -\begin{bmatrix} 0 & a_{12} & \cdots & a_{1n} \\ \vdots & \ddots & \ddots & \vdots \\ \vdots & & \ddots & a_{n-1,n} \\ 0 & \cdots & \cdots & 0 \end{bmatrix}, \tag{1.29}$$

$$L \equiv D^{-1}C_L, \quad U = D^{-1}C_U. \tag{1.30}$$

利用以上记号,矩阵 A 显然有分裂形式

$$A = D - C_L - C_U = D\left(I - L - U\right). \tag{1.31}$$

更一般地,对于分块形式采用同样的记号,便可相仿的建立块迭代格式.

设 A 的分块形式为

$$A=\begin{bmatrix} A_{11} & A_{12} & \cdots & A_{1m} \\ A_{21} & A_{22} & \cdots & A_{2m} \\ \cdots & \cdots & \cdots & \cdots \\ A_{m1} & A_{m2} & \cdots & A_{mm} \end{bmatrix}, \tag{1.32}$$

其中对角块

$$A_{ii}\in\mathbb{C}^{n_i\times n_i},\quad i=1,\cdots,m$$

均是非奇异主子矩阵,

$$n_1+\cdots+n_m=n.$$

相伴地,

$$D\equiv\operatorname{diag}(A_{11},\cdots,A_{mm}), \tag{1.33}$$

$$C_L\equiv-\begin{bmatrix} 0 & \cdots & \cdots & 0 \\ A_{21} & \ddots & & \vdots \\ \vdots & \ddots & \ddots & \vdots \\ A_{m1} & \cdots & A_{m,m-1} & 0 \end{bmatrix}, \tag{1.34}$$

$$C_U\equiv-\begin{bmatrix} 0 & A_{12} & \cdots & A_{1m} \\ \vdots & \ddots & \ddots & \vdots \\ \vdots & & \ddots & A_{m-1,m} \\ 0 & \cdots & \cdots & 0 \end{bmatrix}, \tag{1.35}$$

$$L\equiv D^{-1}C_L,\quad U\equiv D^{-1}C_U. \tag{1.36}$$

仍然,

$$A=D-C_L-C_U=D(I-L-U). \tag{1.37}$$

相应地,方程组(1.1)的分块形式如下:

$$\begin{bmatrix} A_{11} & A_{12} & \cdots & A_{1m} \\ A_{21} & A_{22} & \cdots & A_{2m} \\ \cdots & \cdots & \cdots & \cdots \\ A_{m1} & A_{m2} & \cdots & A_{mm} \end{bmatrix} \begin{bmatrix} x^{(1)} \\ x^{(2)} \\ \vdots \\ x^{(m)} \end{bmatrix} = \begin{bmatrix} b^{(1)} \\ b^{(2)} \\ \vdots \\ b^{(m)} \end{bmatrix} \tag{1.38}$$

采用分块形式的迭代格式称为**块迭代**.相对地,采用非分块形式的迭代格式称为**点迭代**.

8.2 Jacobi 迭代矩阵

8.2.1 定义 设 $A=[a_{ij}]\in\mathbb{C}^{n\times n}$ 分裂为

$$\begin{aligned} &A=M_J-N_J;\\ &M_J=D,N_J=C_L+C_U \end{aligned} \tag{2.1}$$

记号 D,C_L,C_U,L,U 的定义见 **8.1.10**.并且设对角矩阵 D 是非奇异的.

迭代格式

$$Dx^{(k+1)}=(C_L+C_U)x^{(k)}+b,\ k=0,1,2,\cdots \tag{2.2}$$

称为求解方程组(1.1)的 **Jacobi 迭代法**.

记其迭代矩阵为 J,

$$\begin{aligned} J&\equiv M_J^{-1}N_J=D^{-1}(C_L+C_U)\\ &=L+U=I-D^{-1}A \end{aligned} \tag{2.3}$$

称为相伴矩阵 A 的 **Jacobi 矩阵**.

点 Jacobi 迭代法的分量形式如下:

$$a_{ii}x_i^{(k+1)} = -\sum_{j=1,j\neq i}^{n} a_{ij}x_j^{(k)} + b_i,\ i=1,\cdots,n;$$
$$k=0,1,2,\cdots \tag{2.4}$$

点 Jacobi 迭代矩阵形如

$$J = -\begin{bmatrix} 0 & \dfrac{a_{12}}{a_{11}} & \cdots & \cdots & \dfrac{a_{1n}}{a_{11}} \\ \dfrac{a_{21}}{a_{22}} & 0 & \ddots & & \vdots \\ \vdots & \ddots & \ddots & \ddots & \vdots \\ \vdots & & \ddots & \ddots & \dfrac{a_{n-1,n}}{a_{n-1,n-1}} \\ \dfrac{a_{n1}}{a_{nn}} & \cdots & \cdots & \dfrac{a_{n,n-1}}{a_{nn}} & 0 \end{bmatrix}. \tag{2.5}$$

8.2.2 定理 设 $A \in \mathbb{R}^{n\times n}$ 是对称正定矩阵,且 $2D-A$ 正定,则

$$\rho(J) = \rho\left(M_J^{-1}N_J\right) < 1, \tag{2.6}$$

Jacobi 迭代法(2.2)收敛.

证 由于 A 为实对称矩阵,$D^* = D$,得

$$M_J^* + N_J = D^* + C_L + C_U$$
$$= D - A + D = 2D - A, \tag{2.7}$$

由此并从定理假设,知 A 和 $M_J^* + N_J$ 都是正定的.现在依 **8.1.7** 的(1),即得(2.6).

再应用 **8.1.3**,推出 Jacobi 迭代法(2.2)收敛. □

8.2.3 定理 设 $A \in \mathbb{R}^{n\times n}$ 是对称矩阵,$2D-A$ 正定,且 Jacobi 迭代法(2.2)收敛,则 A 是正定的.

证 一方面,因成立(2.7),故

$$M_J^* + N_J = 2D - A$$

是正定的.另一方面,依 **8.1.3**,有 $\rho\left(M_J^{-1}N_J\right) < 1$.因此,依 **8.1.7** 的(2), A 是正定的. □

8.2.4 定理 设 $A = [a_{ij}] \in \mathbb{R}^{n\times n}$, $a_{ii} \neq 0$, $i = 1,\cdots,n$;记

$$D \equiv \mathrm{diag}\left(a_{11},\cdots,a_{nn}\right),\ C \equiv D - A,\ J \equiv D^{-1}C. \tag{2.8}$$

J 是相伴 A 的 Jacobi 矩阵.则 A 为 H-矩阵的充分必要条件条件是

$$\rho\left(|J|\right) < 1. \tag{2.9}$$

证 A 的比较矩阵

$$\mathscr{M}\left(A\right) = |D| - |C| = |D|\left(I - |J|\right),$$

如果 $\rho\left(|J|\right) < 1$,那么 $I - |J|$ 可逆,且

$$\left(I - |J|\right)^{-1} = \sum_{k=0}^{\infty} |J|^k \geq 0.$$

于是,$\mathscr{M}\left(A\right)$ 可逆,且

$$\mathscr{M}\left(A\right)^{-1} = \left(I - |J|\right)^{-1}|D|^{-1} \geq 0.$$

依 **4.7.2**, A 是 H-矩阵.

反之,如果 A 是 H-矩阵,那么 $\mathscr{M}\left(A\right) \geq 0$,从而 $I - |J|$ 可逆,且

$$\left(I - |J|\right)^{-1} = \mathscr{M}\left(A\right)^{-1}|D| \geq 0.$$

利用

$$\left(I - |J|\right)\left(I + |J| + \cdots + |J|^k\right) = I - |J|^{k+1},$$

推出

$$\sum_{i=0}^{k} |J|^i = \left(I - |J|\right)^{-1}\left(I - |J|^{k+1}\right) \leq \left(I - |J|\right)^{-1},\ k = 1,2,\cdots$$

因此 $\sum_{k=0}^{\infty} |J|^k$ 收敛.依 **2.1.23**, $\rho\left(|J|\right) < 1$. □

8.2.5 定理 设 $A=[a_{ij}]\in\mathbb{R}^{n\times n}$ 是严格对角优势或不可约对角优势矩阵,则

$$\rho(J)=\rho\left(M_J^{-1}N_J\right)<1,$$

Jacobi 迭代法(2.2)收敛.

证 依 **2.2.3**, A 是非奇异的,并因而

$$a_{ii}\neq 0, i=1,\cdots,n.$$

点 Jacobi 迭代矩阵 $J=L+U$ 形如(2.5),即 J 的元素为

$$(J)_{ij}=\begin{cases}0, & i=j,\\ -a_{ij}/a_{ii}, & i\neq j,\end{cases}\quad i,j=1,\cdots,n.$$

因此,依 **2.2.2**,当 A 严格对角优势时,

$$\sum_{j=1}^{n}\left|(J)_{ij}\right|=\sum_{j=1,j\neq i}^{n}\left|\frac{a_{ij}}{a_{ii}}\right|<1,\quad i=1,\cdots,n.$$

从而

$$\rho(J)=\|J\|_\infty<1.$$

当 A 不可约对角优势时,

$$\sum_{j=1}^{n}\left|(J)_{ij}\right|=\sum_{j=1,j\neq i}^{n}\left|\frac{a_{ij}}{a_{ii}}\right|\le 1,\quad i=1,\cdots,n.$$

并且至少对一个 i 成立严格不等号,这样,一方面 $\rho(J)=\|J\|_\infty\le 1$;另一方面 $I-J$ 显然也是不可约对角优势的,故 $\det(I-J)\neq 0$,表明 $\rho(J)\neq 1$,从而必有 $\rho(J)<1$.

再应用 **8.1.3**,推出 Jacobi 迭代法(2.2)收敛. □

此定理的另一证法见 **8.6.4**.

8.2.6 定义 对于方程组(1.1)的分块形式(1.38)的迭代格式

$$A_{ii}\left(x^{(i)}\right)^{(k+1)}=-\sum_{j=1,j\neq i}^{m}A_{ij}\left(x^{(j)}\right)^{(k)}+b^{(i)},\ i=1,\cdots,m,$$

$$k=0,1,2,\cdots \quad (2.10)$$

称为求解块方程组(1.38)的**块 Jacobi 迭代法**.

采用分块记号(1.32)~(1.36),迭代格式(2.10)的矩阵形式仍如同(2.2).其迭代矩阵称为相伴 A 的分块形式(1.32)的**块 Jacobi 矩阵**,形式上仍如同(2.3).

8.3 Gauss-Seidel 迭代矩阵

8.3.1 定义 设 $A=[a_{ij}]\in\mathbb{C}^{n\times n}$ 分裂为

$$A=M_G-N_G,$$
$$M_G=D-C_L, N_J=C_U, \quad (3.1)$$

记号 D,C_L,C_U,L,U 的定义见 **8.1.10**.并设矩阵 M_G 是非奇异的.

迭代格式

$$(D-C_L)x^{(k+1)}=C_Ux^{(k)}+b,\ k=0,1,2,\cdots \quad (3.2)$$

称为求解方程组(1.1)的 **Gauss-Seidel 迭代法**.

记其迭代矩阵为 $\mathscr{L}_1$,

$$\mathscr{L}_1\equiv M_G^{-1}N_G=(D-C_L)^{-1}C_U=(I-L)^{-1}U \quad (3.3)$$

称为相伴矩阵 A 的 **Gauss-Seidel 矩阵**.

点 Gauss-Seidel 迭代法的分量形式如下:

$$a_{ii}x_i^{(k+1)}=-\sum_{j=1}^{i-1}a_{ij}x_j^{(k+1)}-\sum_{j=i+1}^{n}a_{ij}x_j^{(k)}+b_i,\ i=1,\cdots,n;$$
$$k=0,1,2,\cdots \quad (3.4)$$

8.3.2 定理 设 $A\in\mathbb{R}^{n\times n}$ 是对称正定矩阵,则

$$\rho(\mathscr{L}_1) = \rho\left(M_G^{-1} N_G\right) < 1, \tag{3.5}$$

Gauss-Seidel 迭代法(3.2)收敛.

证 见 **8.4.2**. □

8.3.3 定理 设 $A = [a_{ij}] \in \mathbb{R}^{n\times n}$ 是严格对角优势或不可约对角优势矩阵,则

$$\rho(\mathscr{L}_1) = \rho\left(M_G^{-1} N_G\right) < 1,$$

Gauss-Seidel 迭代法(3.2)收敛.

证 依 **2.2.3**, A 是非奇异的,并因而

$$a_{ii} \neq 0, i = 1, \cdots, n,$$

因此 $\mathscr{L}_1$ 是有意义的.设 λ 是 $\mathscr{L}_1 = (I - L)^{-1} U$ 的任一特征值,而 $x \neq 0$ 是相应 λ 的特征向量,则 λ 和 x 满足

$$\mathscr{L}_1 x = (I - L)^{-1} Ux = \lambda x.$$

等价地,

$$(\lambda I - \lambda L - U)x = 0. \tag{3.6}$$

现在证明必有 $|\lambda| < 1$.利用反证法.

如果 $|\lambda| \geq 1$,那么在 A 严格对角优势的条件下,

$$|\lambda| \sum_{j=1}^{i-1} \left|\frac{a_{ij}}{a_{ii}}\right| + \sum_{j=i+1}^{n} \left|\frac{a_{ij}}{a_{ii}}\right| \leq |\lambda| \sum_{j=1, j\neq i}^{n} \left|\frac{a_{ij}}{a_{ii}}\right| < |\lambda|, \ i = 1, \cdots, n.$$

在 A 不可约对角优势的条件下,

$$|\lambda| \sum_{j=1}^{i-1} \left|\frac{a_{ij}}{a_{ii}}\right| + \sum_{j=i+1}^{n} \left|\frac{a_{ij}}{a_{ii}}\right| \leq |\lambda| \sum_{j=1, j\neq i}^{n} \left|\frac{a_{ij}}{a_{ii}}\right| \leq |\lambda|, \ i = 1, \cdots, n.$$

并且关于最后的不等号至少对一个 i 成立严格不等号.这表明在两种条件下,方程(3.6)的系数矩阵也是严格对角优势或不可约对角优势的,因而是非奇异的.于是,方程(3.6)只有零解,与 $x \neq 0$ 矛盾.

因此,$\mathscr{L}_1$ 的所有特征值的模均小于1,从而 $\rho(\mathscr{L}_1) < 1$.

再应用 **8.1.3**,推出 Gauss-Seidel 迭代法(3.2)收敛. □

此定理的另一证法见 **8.6.4.**

8.3.4 定理 设 $A=[a_{ij}]\in\mathbb{R}^{n\times n}$, $a_{ii}\neq 0$, $i=1,\cdots,n$, 相伴矩阵 A 的 Jacobi 矩阵 $J=L+U\geq 0$, 则下列四个关系式必有且只有一个成立:

(1) $\rho(J)=\rho(\mathscr{L}_1)=0$.

(2) $0<\rho(\mathscr{L}_1)<\rho(J)<1$.

(3) $\rho(J)=\rho(\mathscr{L}_1)=1$.

(4) $\rho(\mathscr{L}_1)>\rho(J)>1$.

因此,Jacobi 矩阵 J 和 Gauss-Seidel 矩阵 $\mathscr{L}_1$ 或同时收敛,或同时发散.

证 由 $J\geq 0$ 知 $L\geq 0$, $U\geq 0$. 注意到 L 是严格下三角矩阵, 必有 $L^n=0$, 得

$$(I-L)^{-1}=I+L+L^2+\cdots+L^{n-1},$$

推出

$$\mathscr{L}_1=(I-L)^{-1}U\geq 0.$$

因此,依 **4.4.2**,存在 $x\in\mathbb{R}^n$, $x\geq 0$ 且 $x\neq 0$, 使得

$$(I-L)^{-1}Ux=\lambda x,\quad \lambda=\rho(\mathscr{L}_1)\geq 0. \tag{3.7}$$

先考虑 J 不可约的情形. 此时,依 **4.2.4**, $\lambda>0$, 而且 $x>0$, 于是 (3.7)等价于

$$(\lambda L+U)x=\lambda x \tag{3.8}$$

和

$$\left(L+\frac{1}{\lambda}U\right)x=x. \tag{3.9}$$

因为

$$\lambda L+U\geq 0,\qquad L+\frac{1}{\lambda}U\geq 0,$$

而且 $x>0$, 依 **4.2.4**, 从(3.8)和(3.9)推出

$$\rho(\lambda L+U)=\lambda, \quad \rho\left(L+\frac{1}{\lambda}U\right)=1. \tag{3.10}$$

另一方面,依 **4.1.5**,

$$\rho(tL+U) \quad \text{和} \quad \rho(L+tU)$$

对于 $t \geq 0$ 是单调递增的,而且还可以证明是严格单调递增的.因此,若 $\rho(J)=1$,则由 $\rho\left(L+\frac{1}{\lambda}U\right)=1$ 推出 $\lambda=1$;反之,若 $\lambda=1$,则有

$$\rho(J)=\rho(\lambda L+U)=\lambda=1.$$

现在设 $0<\rho(J)<1$.因

$$\rho(L+tU)\big|_{t=1}=\rho(J)<1,$$

又 $\rho\left(L+\frac{1}{\lambda}U\right)=1$,故从 $\rho(L+tU)$ 关于 t 的单调性推出

$$\frac{1}{\lambda}>1, \text{即有} 0<\lambda<1.$$

然而 $\rho(tL+U)$ 关于 t 也严格单调递增,且 $\rho(tL+U)\big|_{t=1}=\rho(J)$,从而推出

$$0<\lambda=\rho(\lambda L+U)=\rho(J)<1.$$

类似地可以证明,若 $\rho(J)>1$,则

$$\lambda \equiv \rho(\mathscr{L}_1)>\rho(J)>1.$$

至于 J 可约的情形,证明从略,可作练习.

最后,(2),(3),(4)在一起,显然蕴涵(1). □

8.3.5 推论 设 Jacobi 矩阵 $J \geq 0$,而且 $0<\rho(J)<1$,则成立渐近收敛速度不等式

$$R(\mathscr{L}_1)>R(J). \tag{3.11}$$

证 从 **8.3.4** 的(2)直接推出. □

8.3.6 定义 对于方程组(1.1)的分块形式(1.38)的迭代格式

$$A_{ii}\left(x^{(i)}\right)^{(k+1)} = -\sum_{j=1}^{i-1} A_{ij}\left(x^{(j)}\right)^{(k+1)} - \sum_{j=i+1}^{m} A_{ij}\left(x^{(j)}\right)^{(k)} + b^{(i)},$$
$$i = 1,\cdots,m, \qquad k = 0,1,2,\cdots \quad (3.12)$$

称为求解块方程组(1.38)的**块 Gauss-Seidel 迭代法**.

采用分块记号(1.32)~(1.36),迭代格式(3.12)的矩阵形式仍如同(3.2).其迭代矩阵称为相伴 A 的分块形式(1.32)的**块 Gauss-Seidel 矩阵**,形式上仍如同(3.3).

8.4 逐次超松弛(SOR)迭代矩阵

8.4.1 定义 设 $A=[a_{ij}]\in\mathbb{C}^{n\times n}$ 分裂为

$$A = M_\omega - N_\omega,$$
$$M_\omega = \frac{1}{\omega}D - C_L, N_\omega = \frac{1-\omega}{\omega}D - C_U, \qquad (4.1)$$

其中 $\omega\in\mathbb{R},\omega\neq 0$,称为**松驰因子**.记号 D,C_L,C_U,L,U 的定义见 **8.1.10**.并设矩阵 M_ω 是非奇异的.

迭代格式

$$(D-\omega C_L)x^{(k+1)} = (\omega C_U + (1-\omega)D)x^{(k)} + \omega b,$$
$$k = 0,1,2,\cdots \qquad (4.2)$$

称为求解方程组(1.1)的**逐次超松弛**(successive overrelaxation)**迭代法**.简称 **SOR 迭代法**.

记其迭代矩阵为 $\mathscr{L}_\omega$,

$$\mathscr{L}_\omega \equiv M_\omega^{-1}N_\omega = (I-\omega L)^{-1}(\omega U + (1-\omega)I), \qquad (4.3)$$

称为相伴矩阵 A 的 **SOR 矩阵**.

点 SOR 迭代法的分量形式如下:

$$a_{ii}x_i^{(k+1)} = a_{ii}x_i^{(k)} + \omega\left\{-\sum_{j=1}^{i-1} a_{ij}x_j^{(k+1)} - \sum_{j=i+1}^{n} a_{ij}x_j^{(k)} + b_i - a_{ii}x_i^{(k)}\right\},$$

$$i = 1,\cdots,n;\quad k = 0,1,2,\cdots \tag{4.4}$$

当 $\omega = 1$ 时, SOR 迭代法就是 Gauss-Seidel 迭代法.

8.4.2 定理 设 $A = [a_{ij}] \in \mathbb{R}^{n\times n}$ 是对称正定矩阵,则当松弛因子 ω 满足

$$0 < \omega < 2$$

时,有

$$\rho(\mathscr{L}_\omega) = \rho(M_\omega^{-1}N_\omega) < 1, \tag{4.5}$$

SOR 迭代法(4.2)收敛.特别,Gauss-Seidel 迭代法(3.2)是收敛的.

证 由于 A 为实对称矩阵, $D^* = D$, $C_L^* = C_U$,得

$$M_\omega^* + N_\omega = \left(\frac{1}{\omega}D - C_L\right)^* + \left(\frac{1-\omega}{\omega}D + C_U\right) = \frac{2-\omega}{\omega}D. \tag{4.6}$$

注意到 A 正定蕴涵 D 正定,故当 $0 < \omega < 2$ 时 $M_\omega^* + N_\omega$ 是正定的.于是,依 **8.1.7** 的(1),即得(4.5).

再应用 **8.1.3**,推出 SOR 迭代法(4.2)收敛. □

8.4.3 定理 设 $A \in \mathbb{R}^{n\times n}$ 是对称矩阵, D 是正定的,且存在 $\omega \in (0,2)$ 使 SOR 迭代法(4.2)收敛,则 A 是正定的.

证 注意到成立(4.6),

$$M_\omega^* + N_\omega = \frac{2-\omega}{\omega}D$$

是正定的.又依 **8.1.3**, $\rho(M_\omega^{-1}N_\omega) < 1$.因此,依 **8.1.7** 的(2), A 是正

定的. $\square$

8.4.4 定理 设 $A=[a_{ij}]\in\mathbb{C}^{n\times n}$, $a_{ii}\neq 0$, $i=1,\cdots,n$,则

$$\rho(\mathscr{L}_\omega)\geq|1-\omega|,\quad \forall\omega\in\mathbb{R}. \tag{4.7}$$

而且仅当 $\mathscr{L}_\omega$ 的所有特征值的模全等于 $|1-\omega|$ 时取等号.

证 考虑 $\mathscr{L}_\omega=(I-\omega L)^{-1}(\omega U+(1-\omega)I)$ 的特征多项式

$$p(\lambda)\equiv\det(\lambda I-\mathscr{L}_\omega).$$

注意到 L 是严格下三角矩阵, $I-\omega L$ 非奇异而且 $\det(I-\omega L)=1$. 于是

$$\begin{aligned}p(\lambda)&=\det(I-\omega L)\det(\lambda I-\mathscr{L}_\omega)\\&=\det((\lambda+\omega-1)I-\omega\lambda L-\omega U).\end{aligned}$$

用 $\lambda_1(\omega),\cdots,\lambda_n(\omega)$ 表示 $\mathscr{L}_\omega$ 的特征值,它们之积与 $p(\lambda)$ 的常数项差一个因子 $(-1)^n$,而 $p(\lambda)$ 的常数项为

$$p(0)=\det((1-\omega)I-\omega U)=(1-\omega)^n,$$

得

$$\prod_{i=1}^{n}\lambda_i(\omega)=(-1)^n(1-\omega)^n.$$

由此有

$$\rho(\mathscr{L}_\omega)=\max_{1\leq i\leq n}|\lambda_i(\omega)|\geq|1-\omega|.$$

而且推出仅当

$$|\lambda_1(\omega)|=\cdots=|\lambda_n(\omega)|=|1-\omega|$$

时 $\rho(\mathscr{L}_\omega)=|1-\omega|$. $\square$

这一定理表明,在讨论 $\mathscr{L}_\omega$ 的收敛性时,对于松弛因子 ω 来说,仅须考虑的取值范围是

$$|1-\omega|<1\qquad\text{或写成}\qquad 0<\omega<2. \tag{4.8}$$

8.4.5 定义 方程组(1.1)的分块形式(1.38)的迭代格式

$$A_{ii}\left(x^{(i)}\right)^{(k+1)} = A_{ii}\left(x^{(i)}\right)^{(k)}$$
$$+\omega\left\{-\sum_{j=1}^{i-1} A_{ij}\left(x^{(j)}\right)^{(k+1)} - \sum_{j=i+1}^{m} A_{ij}\left(x^{(j)}\right)^{(k)} + b^{(i)} - A_{ii}\left(x^{(i)}\right)^{(k)}\right\},$$
$$i = 1,\cdots,m, \quad k = 0,1,2,\cdots \tag{4.9}$$

称为求解块方程组(1.38)的**块 SOR 迭代法**.

采用分块记号(1.32)~(1.36),迭代格式(4.9)的矩阵形式仍如同(4.2).其迭代矩阵称为相伴 A 的分块形式(1.32)的**块 SOR 矩阵**,形式上仍如同(4.3).

8.4.6 定理 设 $A \in \mathbb{R}^{n\times n}$ 是对称矩阵,采用分块记号(1.32)~(1.36),且设 D 是正定的,则块 SOR 迭代法(4.9)收敛的充分必要条件是 A 为正定的且

$$0 < \omega < 2.$$

此定理是 **8.4.2** 和 **8.4.3** 的直接推论.

8.5 对称逐次超松弛(SSOR)迭代矩阵

8.5.1 定义 设 $A \in \mathbb{C}^{n\times n}$ 分裂为

$$A = M_s - N_s,$$

$$M_s = \frac{1}{\omega(2-\omega)}\left(D - \omega\left(C_L + C_U\right) + \omega^2 C_L D^{-1} C_U\right)$$
$$= \frac{1}{\omega(2-\omega)}\left(D - \omega C_L\right) D^{-1}\left(D - \omega C_U\right),$$

$$N_s = \frac{1}{\omega(2-\omega)}\Big((1-\omega)^2 D + \omega(1-\omega)(C_L + C_U)$$
$$+ \omega^2 C_L D^{-1} C_U\Big)$$
$$= \frac{1}{\omega(2-\omega)}((1-\omega)D + \omega C_L)D^{-1}((1-\omega)D + \omega C_U), \tag{5.1}$$

其中 $\omega \in \mathbb{R}, \omega \neq 0$,称为**松弛因子**;记号 D, C_L, C_U, L, U 的定义见 **8.1.10**.并设矩阵 M_s 是非奇异的.

迭代格式

$$(D - \omega C_L)x^{\left(k+\frac{1}{2}\right)} = ((1-\omega)D + \omega C_U)x^{(k)} + \omega b,$$
$$(D - \omega C_U)x^{(k+1)} = ((1-\omega)D + \omega C_L)x^{\left(k+\frac{1}{2}\right)} + \omega b,$$
$$k = 0,1,2,\cdots \tag{5.2}$$

称为求解方程组(1.1)的**对称逐次超松弛**(symmetric successive overrelaxation)**迭代法**, 简称 **SSOR 迭代法**.

记其迭代矩阵为 $\mathscr{P}_\omega$,

$$\mathscr{P}_\omega \equiv M_s^{-1} N_s = \mathscr{U}_\omega \mathscr{L}_\omega;$$
$$\mathscr{U}_\omega = (I - \omega U)^{-1}(\omega L + (1-\omega)I),$$
$$\mathscr{L}_\omega = (I - \omega L)^{-1}(\omega U + (1-\omega)I), \tag{5.3}$$

其中 $\mathscr{L}_\omega$ 就是(4.3)给出的 SOR 迭代矩阵,$\mathscr{P}_\omega$ 称为相伴矩阵 A 的 **SSOR 矩阵**.

8.5.2 定理 设 $A \in \mathbb{R}^{n\times n}$ 是对称正定矩阵,则当 $0 < \omega < 2$ 时,

$$\rho(\mathscr{P}_\omega) = \rho(M_s^{-1} N_s) < 1, \tag{5.4}$$

SSOR 迭代法(5.2)收敛.

证 由于 A 为实对称矩阵,

$$D^* = D, \quad C_L^* = C_U,$$

得

$$M_s^* + N_s = \frac{1}{\omega(2-\omega)}\big((D-\omega C_L)D^{-1}(D-\omega C_U)$$
$$+((1-\omega)D+\omega C_L)D^{-1}((1-\omega)D+\omega C_U)\big). \tag{5.5}$$

注意到 A 正定蕴涵 D 正定和 $(D-\omega C_L)^* = D-\omega C_U$ 非奇异,容易看出当 $0<\omega<2$ 时 $M_s^*+N_s$ 是正定的.于是,依 **8.1.7** 的(1),即得(5.4).

再应用 **8.1.3**,推出 SSOR 迭代法(5.2)收敛. □

8.5.3 定理 设 $A\in\mathbb{R}^{n\times n}$ 是对称矩阵, D 是正定的,且存在 $\omega\in(0,2)$ 使 SSOR 迭代法(5.2)收敛,则 A 是正定的.

证 注意到成立(5.5), $M_s^*+N_s$ 是正定的.又依 **8.1.3**,有 $\rho(M_s^{-1}N_s)<1$.因此,依 **8.1.7** 的(2), A 是正定的. □

8.6 加速超松弛和对称加速超松弛迭代矩阵

8.6.1 定义 设 $A\in\mathbb{C}^{n\times n}$ 分裂为

$$A = M_{\gamma,\omega} - N_{\gamma,\omega};$$
$$M_{\gamma,\omega} = \frac{1}{\omega}D - \frac{\gamma}{\omega}C_L,$$
$$N_{\gamma,\omega} = \frac{1-\omega}{\omega}D + \frac{\omega-\gamma}{\omega}C_L + C_U, \tag{6.1}$$

其中 $\gamma,\omega\in\mathbb{R}, \omega\neq 0$,称为**松弛因子**;记号 D, C_L, C_U, L, U 的定义见 **8.1.10**.并设矩阵 $M_{\gamma,\omega}$ 是非奇异的.

迭代格式

$$(D-\gamma C_L)x^{(k+1)}=((1-\omega)D+(\omega-\gamma)C_L+\omega C_U)x^{(k)}+\omega b,$$
$$k=0,1,2,\cdots \qquad (6.2)$$

称为求解方程组(1.1)的**加速超松弛**(accelerated overrelaxation)**迭代法**, 简称 **AOR 迭代法**.

记其迭代矩阵为 $\mathscr{L}_{\gamma,\omega}(A)$,

$$\mathscr{L}_{\gamma,\omega}(A)\equiv M_{\gamma,\omega}^{-1}N_{\gamma,\omega}$$
$$=(D-\gamma C_L)^{-1}((1-\omega)D+(\omega-\gamma)C_L+\omega C_U) \qquad (6.3)$$

称为相伴矩阵 A 的 **AOR 矩阵**.

迭代格式

$$(D-\gamma C_L)x^{\left(k+\frac{1}{2}\right)}=((1-\omega)D+(\omega-\gamma)C_L+\omega C_U)x^{(k)}+\omega b,$$
$$(D-\gamma C_U)x^{(k+1)}=((1-\omega)D+(\omega-\gamma)C_U+\omega C_L)x^{\left(k+\frac{1}{2}\right)}+\omega b,$$
$$k=0,1,2,\cdots \qquad (6.4)$$

称为求解方程组(1.1)的**对称加速超松弛**(symmetric accelerated overrelaxation)**迭代法**, 简称 **SAOR 迭代法**.

记其迭代矩阵为 $\mathscr{S}_{\gamma,\omega}(A)$,

$$\mathscr{S}_{\gamma,\omega}(A)\equiv \mathscr{U}_{\gamma,\omega}(A)\mathscr{L}_{\gamma,\omega}(A);$$
$$\mathscr{U}_{\gamma,\omega}(A)=(D-\gamma C_U)^{-1}((1-\omega)D+(\omega-\gamma)C_U+\omega C_L), \qquad (6.5)$$

其中 $\mathscr{L}_{\gamma,\omega}(A)$ 是AOR矩阵(6.3). $\mathscr{S}_{\gamma,\omega}(A)$ 称为相伴矩阵 A 的**SAOR矩阵**.

AOR 迭代法是一类广泛的迭代法.特别,

当 $\gamma=0,\omega=1$ 时, AOR 法就是 Jacobi 迭代法.

当 $\gamma=\omega=1$ 时, AOR 法就是 Gauss-Seidel 迭代法.

当 $\gamma=\omega$ 时, AOR 法就是 SOR 迭代法,SAOR 法就是 SSOR 迭代法.

8.6.2 定义 设 $A=[a_{ij}]\in\mathbb{R}^{n\times n}$,集合

$$\Omega(A)\equiv\left\{B=[b_{ij}]\in\mathbb{R}^{n\times n}:|b_{ij}|=|a_{ij}|;i,j=1,\cdots,n\right\} \qquad (6.6)$$

称为 A 的**等模矩阵集合**.

8.6.3 定理 设 $A=[a_{ij}]\in\mathbb{R}^{n\times n}$, $a_{ii}\neq 0$, $i=1,\cdots,n$; J 是相伴 A 的 Jacobi 迭代矩阵; $0\le\gamma\le\omega$,则下列三个条件等价:

(1) A 是 H-矩阵.

(2) 对于任意的 $G\in\Omega(A)$ 和 $\omega\in\left(0,\dfrac{2}{1+\rho(|J|)}\right)$,有

$$\rho(\mathscr{L}_{\gamma,\omega}(G))<1.$$

(3) 对于任意的 $G\in\Omega(A)$ 和 $\omega\in\left(0,\dfrac{2}{1+\rho(|J|)}\right)$,有

$$\rho(\mathscr{P}_{\gamma,\omega}(G))<1.$$

证 记

$$D\equiv\operatorname{diag}(a_{11},\cdots,a_{nn}),\qquad C\equiv D-A,$$

则 $J=D^{-1}C$.

先考虑 A 为不可约矩阵的情形.

先证(1)蕴涵(2)和(3).此时,A 是 H-矩阵.从下述证明可以看出不妨假定 $G=A$.

因 A 不可约,故 $|J|=|D^{-1}||C|$ 是非负不可约矩阵.依 **4.2.4**,存在 $x=(x_1,\cdots,x_n)^{\mathrm{T}}\in\mathbb{R}^n$, $x>0$,使得

$$|D^{-1}||C|x=\rho(|J|)x. \tag{6.7}$$

令

$$Q\equiv\operatorname{diag}(x_1,\cdots,x_n),\qquad \tilde{A}\equiv[\tilde{a}_{ij}]=AQ.$$

注意到

$$\tilde{D}\equiv\operatorname{diag}(\tilde{a}_{11},\cdots,\tilde{a}_{nn})=DQ$$

和

$$\tilde{C}\equiv\tilde{D}-\tilde{A}=CQ,$$

从(6.7)得

$$\sum_{j=1,j\neq i}^{n}\left|\tilde{a}_{ij}\right| = \rho(|J|)\left|\tilde{a}_{ii}\right|, \quad i=1,\cdots,n. \tag{6.8}$$

另一方面,

$$Q^{-1}\mathscr{L}_{\gamma,\omega}(A)Q = \mathscr{L}_{\gamma,\omega}(\tilde{A}) = \tilde{M}^{-1}\tilde{N}, \tag{6.9}$$

其中

$$\tilde{M} = \tilde{D} - \gamma\tilde{C}_L,$$
$$\tilde{N} = (1-\omega)\tilde{D} + (\omega-\gamma)\tilde{C}_L + \omega\tilde{C}_U;$$
$$\tilde{C}_L = C_L Q, \quad \tilde{C}_U = C_U Q.$$

依定理假设,有

$$0 \le \gamma < \frac{2}{1+\rho(|J|)}.$$

又依 **8.2.4**, $\rho(|J|) < 1$.

于是从(6.8),有

$$\begin{aligned}
\left|\tilde{a}_{ii}\right| - \gamma\sum_{j=1}^{i-1}\left|\tilde{a}_{ij}\right| &\ge \left|\tilde{a}_{ii}\right| - \frac{2}{1+\rho(|J|)}\sum_{j=1}^{i-1}\left|\tilde{a}_{ij}\right| \\
&= \frac{1}{1+\rho(|J|)}\left((1+\rho(|J|))\left|\tilde{a}_{ii}\right| - 2\sum_{j=1}^{i-1}\left|\tilde{a}_{ij}\right|\right) \\
&> \frac{1}{1+\rho(|J|)}\left(2\sum_{j=1,j\neq i}^{n}\left|\tilde{a}_{ij}\right| - 2\sum_{j=1}^{i-1}\left|\tilde{a}_{ij}\right|\right) \\
&= \frac{2}{1+\rho(|J|)}\sum_{j=i+1}^{n}\left|\tilde{a}_{ij}\right| \ge 0, \quad i=1,\cdots,n.
\end{aligned}$$

这表明 $\tilde{M}$ 是严格对角优势矩阵.

利用 **8.1.8** 及(6.9),

$$\rho(\mathscr{L}_{\gamma,\omega}(A)) = \rho(\mathscr{L}_{\gamma,\omega}(\tilde{A}))$$

$$\le \left\| \mathscr{L}_{\gamma,\omega}\left(\tilde{A}\right) \right\|_{\infty} = \left\| \tilde{M}^{-1}\tilde{N} \right\|_{\infty}$$

$$\le \max_{1\le i\le n} \frac{|1-\omega||\tilde{a}_{ii}| + (\omega-\gamma)\sum_{j=1}^{i-1}|\tilde{a}_{ij}| + \omega\sum_{j=i+1}^{n}|\tilde{a}_{ij}|}{|\tilde{a}_{ii}| - \gamma\sum_{j=1}^{i-1}|\tilde{a}_{ij}|}$$

$$\le \max_{1\le i\le n} \frac{\left(|1-\omega| + \omega\rho(|J|)\right)|\tilde{a}_{ii}| - \gamma\sum_{j=1}^{i-1}|\tilde{a}_{ij}|}{|\tilde{a}_{ii}| - \gamma\sum_{j=1}^{i-1}|\tilde{a}_{ij}|}$$

$$\le |1-\omega| + \omega\rho(|J|) < 1. \tag{6.10}$$

最后两个不等号基于:当正数 a,b,c 满足 $a>c$ 和 $a\ge b$ 时,成立

$$\frac{b-c}{a-c} \le \frac{b}{a}.$$

至此证明了在条件 $0<\omega<\dfrac{2}{1+\rho(|J|)}$ 下有 $|1-\omega|+\omega\rho(|J|)<1$.因此(1)成立时(2)必成立.

相仿可证

$$\rho\left(\mathscr{U}_{\gamma,\omega}(A)\right) \le \left\|\mathscr{U}_{\gamma,\omega}\left(\tilde{A}\right)\right\|_{\infty} \le |1-\omega| + \omega\rho(|J|) < 1. \tag{6.11}$$

综合(6.10)和(6.11),注意到

$$\mathscr{P}_{\gamma,\omega}\left(\tilde{A}\right) = Q^{-1}\mathscr{P}_{\gamma,\omega}(A)Q,$$

得

$$\begin{aligned}\rho\left(\mathscr{P}_{\gamma,\omega}(A)\right) = \rho\left(\mathscr{P}_{\gamma,\omega}\left(\tilde{A}\right)\right) &\le \left\|\mathscr{P}_{\gamma,\omega}\left(\tilde{A}\right)\right\|_{\infty} \\ &= \left\|\mathscr{U}_{\gamma,\omega}(A)\mathscr{L}_{\gamma,\omega}(A)\right\|_{\infty} \\ &\le \left(|1-\omega| + \omega\rho(|J|)\right)^2 < 1,\end{aligned} \tag{6.12}$$

因此(1)成立时(3)必成立.

现在,反过来证明(1)不成立则(2)和(3)必不成立.此时,A 不是 H-矩阵.取 $G=\mathscr{M}(A)$ 即 A 的比较矩阵,则 G 是 L-矩阵但不是 H-矩阵.依 **8.2.4**,必有 $\rho(|J|)\geq 1$.因此,从(6.8)得

$$|\tilde{a}_{ii}|\leq \sum_{j=1,j\neq i}^{n}|\tilde{a}_{ij}|, \quad i=1,\cdots,n. \tag{6.13}$$

由于

$$|\tilde{a}_{ii}|>0,\ i=1,\cdots,n,$$

故可取充分小的 $\gamma>0$,使得

$$|\tilde{a}_{ii}|-\gamma\sum_{j=1}^{i-1}|\tilde{a}_{ij}|>0, \quad |\tilde{a}_{ii}|-\gamma\sum_{j=i+1}^{n}|\tilde{a}_{ij}|>0,$$
$$i=1,\cdots,n. \tag{6.14}$$

令

$$M=|\tilde{D}|-\gamma|\tilde{C}_L|,\ N=(1-\omega)|\tilde{D}|+(\omega-\gamma)|\tilde{C}_L|+\omega|\tilde{C}_U|,$$

而且

$$l_{ij}\equiv e_i^T M^{-1}Ne_j=e_i^T\mathscr{L}_{\gamma,\omega}(\mathscr{M}(\tilde{A}))e_j, \quad i,j=1,\cdots,n,$$

其中 $\mathscr{M}(\tilde{A})$ 是 $\tilde{A}$ 的比较矩阵.依 **8.1.9**,并利用(6.13)和(6.14),对于任意的

$$\omega\in\left(\gamma,\frac{2}{1+\rho(|J|)}\right),$$

有

$$\begin{aligned}\rho(\mathscr{L}_{\gamma,\omega}(G))&=\rho(\mathscr{L}_{\gamma,\omega}(\mathscr{M}(A)))\\&=\rho(\mathscr{L}_{\gamma,\omega}(\mathscr{M}(\tilde{A})))=\rho(M^{-1}N)\\&\geq\min_{1\leq i\leq n}\sum_{j=1}^{n}l_{ij}\end{aligned}$$

$$\geq \min_{1\leq i\leq n} \frac{(1-\omega)|\tilde{a}_{ii}| + (\omega-\gamma)\sum_{j=1}^{i-1}|\tilde{a}_{ij}| + \omega\sum_{j=i+1}^{n}|\tilde{a}_{ij}|}{|\tilde{a}_{ii}| - \gamma\sum_{j=1}^{i-1}|\tilde{a}_{ij}|} \geq 1. \qquad (6.15)$$

这样,证明了条件(1)不成立时,存在 $G \in \Omega(A)$,满足

$$0 \leq \gamma \leq \omega < \frac{2}{1+\rho(|J|)}.$$

但 $\rho(\mathscr{L}_{\gamma,\omega}(A)) \geq 1$,即条件(2)不成立.

相仿,令

$$u_{ij} \equiv e_i^{\mathrm{T}} \mathscr{U}_{\gamma,\omega}\left(\mathscr{M}\left(\tilde{A}\right)\right) e_j, \quad i,j = 1,\cdots,n,$$

可证

$$\min_{1\leq i\leq n} \sum_{j=1}^{n} u_{ij} \geq 1. \qquad (6.16)$$

再记

$$\beta_{ij} \equiv e_i^{\mathrm{T}} \mathscr{U}_{\gamma,\omega}\left(\mathscr{M}\left(\tilde{A}\right)\right) \mathscr{L}_{\gamma,\omega}\left(\mathscr{M}\left(\tilde{A}\right)\right) e_j, \quad i,j = 1,\cdots,n.$$

依 **8.1.9**,并利用(6.15)和(6.16),得

$$\begin{aligned}
\rho(\mathscr{P}_{\gamma,\omega}(G)) &= \rho(\mathscr{P}_{\gamma,\omega}(\mathscr{M}(A))) = \rho\left(\mathscr{P}_{\gamma,\omega}\left(\mathscr{M}\left(\tilde{A}\right)\right)\right) \\
&\geq \min_{1\leq i\leq n} \sum_{j=1}^{n} \beta_{ij} = \min_{1\leq i\leq n} \sum_{j=1}^{n} \sum_{k=1}^{n} u_{ik} l_{kj} \\
&= \min_{1\leq i\leq n} \sum_{k=1}^{n} u_{ik} \sum_{j=1}^{n} l_{kj} \\
&\geq \min_{1\leq i\leq n} \sum_{k=1}^{n} u_{ik} \left(\min_{1\leq k\leq n} \sum_{j=1}^{n} l_{kj} \right) \\
&\geq \left(\min_{1\leq i\leq n} \sum_{k=1}^{n} u_{ik} \right) \left(\min_{1\leq k\leq n} \sum_{j=1}^{n} l_{kj} \right) \geq 1.
\end{aligned}$$

这表明(1)不成立时(3)必不成立.

至此,对于 A 不可约的情形证明了定理.

下面考虑 A 可约的情形.采用通常的处理方法:将 A 的所有零元素替换为 ε,所得矩阵是不可约的,记之为 A_ε,这样便可把有关不可约矩阵的结论应用于 A_ε;然后,利用矩阵的连续性质,从 A_ε 引出 A 的相应结论.在这里,套用如此模式并不困难,因而只以推证(1)蕴涵(2)和(3)为示范,推证(1)不成立时(2)和(3)必不成立则留作练习.

由于 A 是 H-矩阵,依 **8.2.4**,有 $\rho(|J|)<1$,其中 $J=D^{-1}C$ 是相伴 A 的 Jacobi 矩阵.又 A 的对角元素均不为零,相伴 A_ε 的 Jacobi 矩阵必形如 $J_\varepsilon=D^{-1}C_\varepsilon$.鉴于特征值连续地依赖于矩阵的元素,当 $\varepsilon\to 0$ 时,有

$$\rho(|J_\varepsilon|)\to\rho(|J|)<1,$$

因此,必存在 $\varepsilon_1>0$,使得当 $0<\varepsilon<\varepsilon_1$ 时,$\rho(|J_\varepsilon|)<1$.这样,仍依 **8.2.4**, A_ε 是 H-矩阵.于是,当

$$0<\varepsilon<\varepsilon_1,\quad 0\le\gamma\le\omega<\frac{2}{1+\rho(|J_\varepsilon|)}$$

时,对 A_ε 仿照(6.10)和(6.12)的推导,有

$$\rho(\mathscr{L}_{\gamma,\omega}(A_\varepsilon))\le|1-\omega|+\omega\rho(|J_\varepsilon|) \tag{6.17}$$

和

$$\rho(\mathscr{S}_{\gamma,\omega}(A_\varepsilon))\le(|1-\omega|+\omega\rho(|J_\varepsilon|))^2. \tag{6.18}$$

注意到对于任意取定的 γ 和 ω, $0\le\gamma\le\omega<\dfrac{2}{1+\rho(|J_\varepsilon|)}$,必存在 $\varepsilon_2>0$, $\varepsilon_2\le\varepsilon_1$ 使得当 $0<\varepsilon<\varepsilon_2$ 时,有

$$0\le\gamma\le\omega<\frac{2}{1+\rho(|J_\varepsilon|)}.$$

从而,当$0<\varepsilon<\varepsilon_2$时,(6.17)和(6.18)成立.进而,在(6.17)和(6.18)中令$\varepsilon\to 0$,得

$$\rho(\mathscr{L}_{\gamma,\omega}(A))\le|1-\omega|+\omega\rho(|J|)<1$$

和

$$\rho(\mathscr{S}_{\gamma,\omega}(A))\le(|1-\omega|+\omega\rho(|J|))^2<1.$$

至此,对于A可约的情形,证明了(1)蕴涵(2)和(3). □

8.6.4 推论 设A是严格对角优势矩阵或不可约弱严格对角优势矩阵,$0\le\gamma\le\omega<\dfrac{2}{1+\rho(|J|)}$,其中$J$是相伴$A$的 Jacobi 矩阵,则对应的 Jacobi,Gauss-seidel,SOR 和 SSOR 迭代法都是收敛的.

证 直接从 **4.7.5** 和 **8.6.3** 推出. □

8.7 矩阵的正则分裂

8.7.1 定义 设$A\in\mathbb{R}^{n\times n}$,$A=M-N$称为矩阵$A$的**正则分裂**,如果

(1) $M\in\mathbb{R}^{n\times n}$是非奇异的且$M^{-1}\ge 0$;

(2) $N\in\mathbb{R}^{n\times n}$是非负的,即$N\ge 0$.

$A=M-N$称为矩阵A的**弱正则分裂**,如果

(1) $M^{-1}\ge 0$;

(2) $G\equiv M^{-1}N\ge 0$.

8.7.2 定理 设$A=M-N$是矩阵A的正则分裂,且$A^{-1}\ge 0$,则

$$\rho(M^{-1}N)=\frac{\rho(A^{-1}N)}{1+\rho(A^{-1}N)}<1, \tag{7.1}$$

迭代格式(1.7)收敛.

证 根据定理的假设,有

$$G \equiv M^{-1}N = (A+N)^{-1}N = (I+F)^{-1}F, \\ F \equiv A^{-1}N. \tag{7.2}$$

而且$G \geq 0, F \geq 0$.依 **4.4.2**,$\rho(G)$是G的特征值,且存在相应的非负特征向量.

另一方面,若x是F的相应于特征值τ的特征向量,则

$$Fx = \tau x, \quad Gx = (I+F)^{-1}Fx = \frac{\tau}{1+\tau}x.$$

这表明x也是G的特征向量,对应的特征值为

$$\mu = \frac{\tau}{1+\tau}. \tag{7.3}$$

反之,若μ是G的特征值,$Gz = (I+F)^{-1}Fz = \mu z$,则

$$Fz = \mu(I+F)z.$$

由此,推出$\mu \neq 1$,从而

$$Fz = \frac{\mu}{1-\mu}z \equiv \tau z.$$

这样,仍然得出特征值关系式(7.3).

现在取x是G的相应于$\rho(G)$的非负特征向量.由于F也是非负矩阵,F对应x的特征值τ必为非负数.显然,对于$\tau \geq 0$,

$$\tau/(1+\tau)$$

是单调递增的,因此选取$\tau = \rho(F)$时(7.3)中的μ在$\lambda(G)$中按模达最大值,即成立

$$\rho(G) = \frac{\rho(F)}{1+\rho(F)}.$$

由此式直接看出$\rho(G) < 1$,从而(7.1)得证,再依 **8.1.3**,迭代格式(1.7)收敛. □

8.7.3 推论 设$A \in \mathbb{R}^{n \times n}$是 Stieltjes 矩阵,$A = M - N$是$A$的正则分裂,其中$N$是实对称的,则

$$\rho\left(M^{-1}N\right)=\frac{\rho(N)\rho\left(A^{-1}\right)}{1+\rho(N)\rho\left(A^{-1}\right)}<1. \tag{7.4}$$

证 依 **4.6.14**,有 $A^{-1}\geq 0$,从而(7.1)成立.因为 A^{-1} 和 N 都是实对称矩阵,依 **3.6.9**,有

$$\rho\left(A^{-1}\right)=\left\|A^{-1}\right\|_2, \qquad \rho(N)=\|N\|_2,$$

所以

$$\rho\left(A^{-1}N\right)\leq\left\|A^{-1}N\right\|_2\leq\left\|A^{-1}\right\|_2\|N\|_2=\rho\left(A^{-1}\right)\rho(N).$$

由此并从(7.1)及 $\tau\geq 0$ 时 $\tau/(1+\tau)$ 的单调性即得(7.4). □

8.7.4 定理 设 $A\in\mathbb{R}^{n\times n}$ 是M-矩阵,M 是从 A 置某些非对角元素为零而得的矩阵,并令 $N\equiv M-A$,则 $A=M-N$ 是 A 的正则分裂,而且 $\rho\left(M^{-1}N\right)<1$.

证 直接从 **4.6.6** 的(1)和 **8.7.2** 推出. □

8.7.5 定理 设 $A\in\mathbb{R}^{n\times n}$, $A^{-1}>0$.又设

$$A=M_1-N_1=M_2-N_2$$

是 A 的两种正则分裂,满足 $N_2\geq N_1\geq 0$ 而且排除等号(即 N_1 和 N_2-N_1 均非零矩阵), 则

$$1>\rho\left(M_2^{-1}N_2\right)>\rho\left(M_1^{-1}N_1\right)>0. \tag{7.5}$$

证 应用(7.1),$\rho\left(M^{-1}N\right)$是关于 $\rho\left(A^{-1}N\right)$单调递增的,因此只须证明

$$\rho\left(A^{-1}N_2\right)>\rho\left(A^{-1}N_1\right)>0. \tag{7.6}$$

因为 $N_2\geq N_1\geq 0$,排除等号,且 $A^{-1}>0$,有

$$A^{-1}N_2\geq A^{-1}N_1\geq 0, \tag{7.7}$$

且排除等号.

先考虑 $A^{-1}N_1$ 是不可约的情形.由(7.6)看出,$A^{-1}N_2$ 从而 $\left(A^{-1}N_2\right)^{\mathrm{T}}$ 也是不可约的.依 **4.2.4**,存在 $x,y\in\mathbb{R}^n$, $x>0,y>0$,

成立

$$\left(A^{-1}N_1\right)x=\rho\left(A^{-1}N_1\right)x$$

和

$$\left(A^{-1}N_2\right)^{\mathrm{T}}y=\rho\left(A^{-1}N_2\right)y \quad 或 \quad y^{\mathrm{T}}\left(A^{-1}N_2\right)=\rho\left(A^{-1}N_2\right)y^{\mathrm{T}}.$$

由此以及(7.7)并排除等号,有

$$\begin{aligned} y^{\mathrm{T}}\left(A^{-1}N_2\right)x &= \rho\left(A^{-1}N_2\right)y^{\mathrm{T}}x \\ &> y^{\mathrm{T}}\left(A^{-1}N_1\right)x=\rho\left(A^{-1}N_1\right)y^{\mathrm{T}}x>0. \end{aligned}$$

此式因 $y^{\mathrm{T}}x\in\mathbb{R}$, $y^{\mathrm{T}}x>0$,故等价于(7.6).

现在先考虑 $A^{-1}N_1$ 是可约的情形.鉴于 $A^{-1}>0$ 且 $N_1\geq 0$,如果 $A^{-1}N_1$ 的 (i,j)-元素为 0,那么 N_1 的第 j 列元素必全为 0.因此必存在置换矩阵 $P\in\mathbb{R}^{n\times n}$,使得

$$P\left(A^{-1}N_1\right)P^{\mathrm{T}}=\begin{bmatrix}0 & B\\ 0 & C\end{bmatrix},$$

其中

$$C\in\mathbb{R}^{r\times r}, C>0, B\in\mathbb{R}^{(n-r)\times r}, B>0.$$

并且由 $N_1\neq 0$ 得知 $1\leq r\leq n-1$.依 **4.2.4**,存在 C 的相应 $\rho(C)$ 的特征向量 $v\in\mathbb{R}^{r}$, $v>0$,成立

$$Cv=\rho(C)v.$$

由于

$$\rho\left(A^{-1}N_1\right)=\rho\left(P\left(A^{-1}N_1\right)P^{\mathrm{T}}\right)=\rho(C)>0,$$

容易看出,可取 $P\left(A^{-1}N_1\right)P^{\mathrm{T}}$ 的相应 $\rho\left(A^{-1}N_1\right)$ 的特征向量为

$$x=\begin{bmatrix}u\\ v\end{bmatrix}, \qquad u=\frac{1}{\rho\left(A^{-1}N_1\right)}Bv,$$

显然, $x>0$.另一方面,从(7.7)有

$$P\left(A^{-1}N_2\right)P^{\mathrm{T}}\geq P\left(A^{-1}N_1\right)P^{\mathrm{T}}\geq 0.$$

注意到

$$\rho\left(A^{-1}N_2\right)=\rho\left(P\left(A^{-1}N_2\right)P^{\mathrm{T}}\right).$$

依 **4.4.2**,存在 $y\in\mathbb{R}^n$, $y\geq 0$, $y\neq 0$,使得

$$\left(P\left(A^{-1}N_2\right)P^{\mathrm{T}}\right)^{\mathrm{T}}y=\rho\left(A^{-1}N_2\right)y$$

或

$$y^{\mathrm{T}}\left(P\left(A^{-1}N_2\right)P^{\mathrm{T}}\right)=\rho\left(A^{-1}N_2\right)y^{\mathrm{T}}.$$

于是

$$\begin{aligned}y^{\mathrm{T}}\left(P\left(A^{-1}N_2\right)P^{\mathrm{T}}\right)x&=\rho\left(A^{-1}N_2\right)y^{\mathrm{T}}x\\&>y^{\mathrm{T}}\left(P\left(A^{-1}N_1\right)P^{\mathrm{T}}\right)x=\rho\left(A^{-1}N_1\right)y^{\mathrm{T}}x>0.\end{aligned}$$

此式仍有 $y^{\mathrm{T}}x>0$,故仍等价于(7.6). □

8.7.6 推论 设 $A\in\mathbb{R}^{n\times n}$, $A^{-1}>0$, $A=M_1-N_1=M_2-N_2$ 是 A 的两种正则分裂,满足 $N_2\geq N_1\geq 0$ 而且排除等号,则存在渐近收敛速度不等式

$$R\left(M_1^{-1}N_1\right)>R\left(M_2^{-1}N_2\right)>0. \tag{7.8}$$

证 从(1.14)和(7.5)直接推出. □

8.7.7 推论 设 $A\in\mathbb{R}^{n\times n}$ 是不可约的 Stieltjes 矩阵,$M_1,M_2\in\mathbb{R}^{n\times n}$ 均由置 A 的某些非对角元素为零而得,而且 $A=M_1-N_1$ 和 $A=M_2-N_2$ 满足 $N_2\geq N_1\geq 0$,排除等号,则

$$0<\rho\left(M_1^{-1}N_1\right)<\rho\left(M_2^{-1}N_2\right)<1.$$

证 从 **4.6.14**,**4.6.4**,**8.7.4** 和 **8.7.5** 直接推出. □

8.7.8 定理 设 $A\in\mathbb{R}^{n\times n}$, $A=I-B$,其中 $B=L+U$ 非负不可约的收敛矩阵,L 和 U 分别是严格下和上三角矩阵,则(4.3)的 SOR 矩阵 $\mathscr{L}_\omega$ 对于所有 $0<\omega\leq 1$ 是收敛的,而且当

$$0<\omega_1<\omega_2\leq 1$$

时,成立

$$0<\rho(\mathscr{L}_{\omega_2})<\rho(\mathscr{L}_{\omega_1})<1, \tag{7.9}$$

并因此

$$R(\mathscr{L}_{\omega_1})<R(\mathscr{L}_{\omega_2}). \tag{7.10}$$

证 定义

$$A=M_\omega-N_\omega,$$

其中依(4.1),

$$M_\omega=\frac{1}{\omega}I-L, \qquad N_\omega=\frac{1-\omega}{\omega}I+U.$$

根据定理假设,显然,对于所有$0<\omega\le 1$来说,$A=M_\omega-N_\omega$是正则分裂;而且

$$A=I-B,\ B\ge 0,\ \rho(B)<1.$$

依 **4.6.5** 的(1)和(3)等价,知$A^{-1}\ge 0$.另一方面,容易看出,当$0<\omega_1<\omega_2\le 1$时,有$0\le N_{\omega_2}\le N_{\omega_1}$,且排除等号.这样,应用 **8.7.5** 和 **8.7.6** 即得(7.9)和(7.10). □

8.8 交替方向隐式迭代(ADI)矩阵

8.8.1 定义 设$A\in\mathbb{C}^{n\times n}$分裂为

$$A=H+V. \tag{8.1}$$

方程组(1.1)即$Ax=b$可以写成如下一对等价方程组

$$\begin{aligned}(H+rI)x&=(rI-V)x+b,\\(V+rI)x&=(rI-H)x+b,\end{aligned}\qquad r\in\mathbb{R},\ r>0. \tag{8.2}$$

引入迭代格式

$$\begin{aligned}(H+r_{k+1}I)x^{\left(k+\frac{1}{2}\right)}&=(r_{k+1}I-V)x^{(k)}+b,\\(V+r_{k+1}I)x^{(k+1)}&=(r_{k+1}I-H)x^{\left(k+\frac{1}{2}\right)}+b,\end{aligned}$$

$$k=0,1,2\cdots \tag{8.3}$$

其中 r_k 称为**加速参数**,(8.3)称为求解方程组(1.1)的 **Peaceman-Rachford 隐式交替方向迭代**(implicit alternating direction iterative) **法**,简称 **ADI 法**.

记其迭代矩阵为 T_r,

$$T_r \equiv (V+rI)^{-1}(rI-H)(H+rI)^{-1}(rI-V) \tag{8.4}$$

称为相伴矩阵 A 的 **Peaceman-Rachford 矩阵**.

交替方向法一词源于形如(8.3)的迭代格式最初是为数值求解偏微分方程问题而设计的.

考虑以如下 Dirichlet 问题为模型问题:

$$\begin{cases} -\Delta u(x,y) = f(x,y), & (x,y)\in \Omega, \\ u(x,y) = g(x,y), & (x,y)\in \partial\Omega, \end{cases} \tag{8.5}$$

其中

$$\Delta u(x,y) \equiv \frac{\partial^2 u(x,y)}{\partial x^2} + \frac{\partial^2 u(x,y)}{\partial y^2},$$

而

$$\Omega \equiv \{(x,y): 0 < x, y < 1\}$$

是单位正方域,$\partial\Omega$ 是 Ω 的边界,f 和 g 是已知函数.

以等步长 $h = 1/m$,通过水平方向(平行于 x 轴)和垂直方向(平行于 y 轴)两组网格线,将正方域 Ω 剖分为若干正方形网格,其结点为

$$(ih, jh), \qquad i, j = 0, 1, \cdots, m.$$

然后,在每个内部结点 (ih, jh)($1 \le i, j \le m-1$)处以中心差商逼近偏导数:

$$\begin{aligned} &\Delta_h u(ih, jh) \\ &\equiv \frac{u((i+1)h, jh) - 2u(ih, jh) + u((i-1)h, jh)}{h^2} \\ &\quad + \frac{u(ih, (j+1)h) - 2u(ih, jh) + u(ih, (j-1)h)}{h^2} \end{aligned} \tag{8.6}$$

这样,问题(8.5)近似地离散成

$$\begin{cases}-\Delta_h u(ih,jh)=f(ih,jh), & (ih,jh)\in\Omega,\\ u(ih,jh)=g(ih,jh), & (ih,jh)\in\partial\Omega.\end{cases} \tag{8.7}$$

对结点加以适当的编号之后,(8.7)可以写成如下形式的线性方程组:

$$A_h u_h = b_h, \tag{8.8}$$

其中 A_h 是 $n\equiv(m-1)^2$ 阶方阵,其每行至多有5个非零元素;u_h 是含 n 个分量的待求向量,其分量是待求的

$$u(ih,jh),\quad \forall(ih,jh)\in\Omega.$$

实际计算时得到的是每个 $u(ih,jh)$ 的近似值;b_h 是由 f 和 g 确定的已知向量.

注意到(8.6)所含的两项,其前一项是沿水平网格线的,后一项是沿垂直网格线的,因此可将 A_h 分裂成相应的两项

$$A_h = H_h + V_h,$$

从而采用形如(8.3)的迭代格式时,其第一个方程是沿水平网格线求解,第二个方程是沿垂直网格线求解,两个方向交替进行.

8.8.2 定理 设 $H,V\in\mathbb{C}^{n\times n}$ 是自伴半正定矩阵,而且两者至少有一个是正定的,则对于任何 $r>0$,由(8.4)确定的 Peaceman-Rachford 矩阵 T_r 是收敛的.

证 只须推证 $\rho(T_r)<1$,考虑

$$\tilde{T}_r\equiv(V+rI)T_r(V+rI)^{-1}.$$

依(8.4),

$$\tilde{T}_r=(rI-H)(H+rI)^{-1}(rI-V)(V+rI)^{-1},$$

于是,由 $\tilde{T}_r$ 相似于 T_r,有

$$\begin{aligned}\rho(T_r)=\rho(\tilde{T}_r)&\le\|\tilde{T}_r\|_2\\ &\le\|(rI-H)(H+rI)^{-1}\|_2\|(rI-V)(V+rI)^{-1}\|_2.\end{aligned} \tag{8.9}$$

根据假设,H 和 V 的特征值都是非负实数.不妨设 H 是正定的,则其特征值还是正数.用 λ_i 表示 H 的特征值,x_i 是相应的特征向量,$i=1,\cdots,n$.从

$$Hx_i=\lambda_i x_i,\quad i=1,\cdots,n,$$

推出

$$(rI-H)x_i=(r-\lambda_i)x_i,\quad i=1,\cdots,n$$

和

$$(H+rI)^{-1}x_i=\frac{1}{r+\lambda_i}x_i,\ i=1,\cdots,n.$$

因而,$(rI-H)(H+rI)^{-1}$ 的特征值为

$$\frac{r-\lambda_i}{r+\lambda_i},\quad i=1,\cdots,n.$$

注意到 $(rI-H)(H+rI)^{-1}$ 是自伴的,而且

$$r>0,\lambda_i>0,\quad i=1,\cdots,n.$$

依 **3.6.9**,有

$$\left\|(rI-H)(H+rI)^{-1}\right\|_2=\max_{1\le i\le n}\left|\frac{r-\lambda_i}{r+\lambda_i}\right|<1. \tag{8.10}$$

类似的论证应用于

$$(rI-V)(V+rI)^{-1}.$$

注意到 V 的特征值可以为零,有

$$\left\|(rI-V)(V+rI)^{-1}\right\|_2\le 1. \tag{8.11}$$

综合(8.9),(8.10)和(8.11),得 $\rho(T_r)<1$. □

9 矩阵函数和函数矩阵

9.1 矩阵和函数

9.1.1 引言 从其余各章可以提取许多形如

$$f:\mathbb{C}^{m\times n}\to\mathbb{C}^{p\times q} \tag{1.1}$$

的映射,它们依赖于 m,n,p,q 的值以及对应关系.例如

(1) $f:\mathbb{C}^{n\times n}\to\mathbb{C}(\equiv\mathbb{C}^{1\times 1}), f(A)=\text{tr}A$ 是 $A\in\mathbb{C}^{n\times n}$ 的迹.

(2) $f:\mathbb{C}^{n\times n}\to\mathbb{C}, f(A)=\det A$ 是 $A\in\mathbb{C}^{n\times n}$ 的行列式.

(3) $f:\mathbb{C}^{m\times n}\to\mathbb{C}^{mn}, f(A)=\text{vec}A$ 是由 $A\in\mathbb{C}^{m\times n}$ 的列连接而成的向量.

(4) $f:\mathbb{C}^{m\times n}\to\mathbb{C}^{m\times m}, f(A)=AA^*$.

(5) $f:\mathbb{C}^{m\times n}\to\mathbb{C}^{n\times n}, f(A)=A^*A$.

(6) $f:\mathbb{C}^{m\times n}\to\mathbb{C}^{\binom{m}{k}\times\binom{n}{k}}$, $f(A)=C_k(A)$是 $A\in\mathbb{C}^{m\times n}$ 的第 k 合成矩阵.

(7) $f:\mathbb{C}^{m\times n}\to\mathbb{C}^{mr\times ns}, f(A)=A\otimes B$ 是矩阵 $A\in\mathbb{C}^{m\times n}$ 和矩阵 $B\in\mathbb{C}^{r\times s}$ 的 Kronecker 积.

(8) $f:\mathbb{C}^{m\times n}\to\mathbb{C}^{m\times n}, f(A)=A\circ A$ 是矩阵 $A\in\mathbb{C}^{m\times n}$ 和其自己的 Hadamard 积.

(9) $f:\mathbb{C}^{n\times n}\to\mathbb{C}^{n\times n}, f(A)=\sum_{i=0}^{m}a_iA^i\,(a_0,\cdots,a_m\in\mathbb{C})$是矩阵 $A\in\mathbb{C}^{n\times n}$ 的多项式.

(10) $f:\mathbb{C}\to\mathbb{C}^{m\times n}$, $f(t)=A(t)\equiv\left[a_{ij}(t)\right]\in\mathbb{C}^{m\times n}$ 是以函数为元素的矩阵.

这些例子包含三种情形:

第一种情形如(1)和(2),值域是数集,f 符合传统的函数概念.

第二种情形如(10),定义域是数集,f 是**矩阵值**(包括**向量值**)**函数**,其矩阵的每个元素是传统的一元复或实变量 t 的函数.

第三种情形如余下的(3)~(9),f 都是矩阵集合到矩阵集合(包括向量集合)的映射.

无论何种情形,现在常一概采用"函数"一词.这样,依定义域来区分:第一和第三两种情形定义域是矩阵集合,f 可称为**矩阵**(的)**函数**;第二种情形定义域是数集,f 确定的是以通常函数为元素的矩阵,可称为**函数**(的)**矩阵**.

不过,上述例子中的 f 都是依已有概念有明确定义的,对它们的讨论似乎并不很依赖这里的函数观念.

新的更一般的一个问题是:如果 A 是一个矩阵,f 是一个通常的一元实或复值函数,那么 $f(A)$ 的意思是需要加以定义的,应当怎样来进行定义呢?集中考虑 $\mathbb{C}^{n\times n}\to\mathbb{C}^{n\times n}$ 的情形.这种情形涉及广泛的自然函数类型,因为它包括着当 f 是多项式时,可以按自然方式定义相伴 f 的多项式矩阵函数.当 f 是解析函数时,其幂级数展开是多项式的自然推广,从而可以建立相伴的 $f(A)$ 的概念.另外,对于不是多项式或者不能用幂级数表示的 f,则须另辟蹊径,并且已利用矩阵 A 的谱及插值多项式给出 $f(A)$ 的一种定义.

存在其它并非基于多项式矩阵函数的关于 $f(A)$ 的有用概念.例如,一种思想是矩阵 $A\in\mathbb{C}^{n\times n}$ 出自连续积分核

$$K(\cdot,\cdot):[a,b]\times[a,b]\to\mathbb{C}$$

的离散化, $A=[a_{ij}]\equiv\left[K\left(x_i,y_j\right)\right]$,并导致考虑核的(逐次)幂

$$K^2(x,y),K^3(x,y),\cdots$$

将它们离散化,即是 Hadamard 幂 $A\circ A, A\circ A\circ A,\cdots$ 更一般的核函数 $f(K(\cdot,\cdot))$,产生所谓 Hadamard 矩阵函数

$$\left[f\left(K\left(x_i,y_j\right)\right)\right]=\left[f\left(a_{ij}\right)\right].$$

反过来的问题是考虑形如

$$f(X)=A$$

的某些类型的非线性矩阵方程.利用标准型,矩阵分解,以及其它工具来求解.求矩阵的平方根($X^2=A$)或对数($\mathrm{e}^X=A$)就是重要的例子.

函数(的)矩阵 $A(t)=\left[a_{ij}(t)\right]$ 称为对于 t 是**连续**的,**可积**的,或**可微**的,如果其所有元素分别是 t 的连续的,可积的,或可微的函数.由此出发,建立了函数矩阵的初等微分学,导出一些值得重视的恒等式和不等式.特别,每个 $a_{ij}(\cdot)$ 是多项式的函数矩阵 $A(\cdot)$,传统上以 λ 为参变量,称 $A(\lambda)$ 为 **λ-矩阵**.有一系列关于 λ-矩阵的基本结果.

更进一步是讨论 $A(t)$ 的多项式矩阵函数 $f(A(t))$,对其导数的直接表示式虽然不难计算,但是因不可交换会有很繁复的形式.因此,有必要考虑 $f(A(t))$ 的导数的一般公式.

9.1.2 定义 设 $A\in\mathbb{C}^{n\times n}$,其谱为

$$\lambda(A)=\left\{\lambda_1,\cdots,\lambda_s:\lambda_i\neq\lambda_j,i\neq j\right\},$$

而且

$$m(t)=\left(t-\lambda_1\right)^{m_1}\cdots\left(t-\lambda_s\right)^{m_s} \tag{1.2}$$

是 A 的最小多项式,其中 $m_k\geq 1$ 是 A 的 λ_k 的指数, $k=1,\cdots,s$.如

果 f 是一元实或复值函数,存在

$$f(\lambda_k), f^{(1)}(\lambda_k), \cdots, f^{(m_k-1)}(\lambda_k),$$

$$k=1,\cdots,s, \tag{1.3}$$

则称 f 为**定义在 A 的谱 $\lambda(A)$ 上的函数**.(1.3)称为 f **在 $\lambda(A)$ 的值**.

显然,每一复系数多项式在任一 $A \in \mathbb{C}^{n\times n}$ 的谱上有定义.特别, A 的最小多项式在 $\lambda(A)$ 上的值全为零.

9.1.3 定理 设 $\lambda_1,\cdots,\lambda_s$ 是不同的实或复数,而 $m_1,\cdots,m_s$ 是正整数,

$$m=m_1+\cdots+m_s.$$

设 f 是一元实或复值函数,存在

$$f^{(j)}(\lambda_k), \quad j=0,1,\cdots,m_k-1;$$

$$k=1,\cdots,s, \tag{1.4}$$

则存在唯一的次数 $\le m-1$ 的多项式 $p(t)$,使得

$$p^{(j)}(\lambda_k)=f^{(j)}(\lambda_k), \quad j=0,1,\cdots,m_k-1;$$

$$k=1,\cdots,s. \tag{1.5}$$

证 对于每个 k, $1\le k\le s$,确定一个次数 $\le m-1$ 的多项式 $p_k(t)$ 如下:

$$p_k(t)=\alpha_k(t)\psi_k(t), \tag{1.6}$$

其中 $\alpha_k(t)$ 形如

$$\alpha_k(t)=\alpha_{k,0}+\alpha_{k,1}(t-\lambda_k)+\cdots+\alpha_{k,m_k-1}(t-\lambda_k)^{m_k-1}, \tag{1.7}$$

$\psi_k(t)$ 取作

$$\psi_k(t)=\begin{cases} 1, & s=1, \\ \prod_{r=1,r\ne k}^{s}(t-\lambda_r)^{m_r}, & s>1. \end{cases} \tag{1.8}$$

容易看出,对于任意的

$$\alpha_{k,0}, \alpha_{k,1}, \cdots, \alpha_{k,m_k-1},$$

$p_k(t)$满足条件

$$p_k(\lambda_i) = p_k^{(1)}(\lambda_i) = \cdots = p_k^{(m_i-1)}(\lambda_i) = 0, \ \forall i \neq k.$$

因此,多项式

$$p(t) \equiv p_1(t) + \cdots + p_s(t) \tag{1.9}$$

满足条件(1.5)的充分必要条件是

$$p_k^{(j)}(\lambda_k) = f^{(j)}(\lambda_k), \quad j = 0,1,\cdots,m_k - 1;$$
$$k = 1,\cdots,s. \tag{1.10}$$

根据(1.6), $p_k(t)$的导数

$$p_k^{(j)}(\lambda_k) = \sum_{i=0}^{j} \binom{j}{i} \alpha_k^{(i)}(\lambda_k) \psi_k^{(j-i)}(\lambda_k),$$
$$j = 0,1,\cdots,m_k - 1; \quad k = 1,\cdots,s.$$

再利用(1.10)和(1.7),有

$$f^{(j)}(\lambda_k) = \sum_{i=0}^{j} \binom{j}{i} i! \alpha_{k,i} \psi_k^{(j-i)}(\lambda_k),$$
$$j = 0,1,\cdots,m_k - 1; \quad k = 1,\cdots,s. \tag{1.11}$$

因为对每个固定的k,

$$\psi_k(\lambda_k) \neq 0,$$

所以由(1.11)从 $j = 0$ 开始递推,可以得到使(1.10)成立的系数$\alpha_{k,0}, \alpha_{k,1}, \cdots, \alpha_{k,m_k-1}$.于是,由(1.9)给出的多项式$p(t)$满足所要求的条件(1.5).

从次数$\leq m-1$的多项式含m个系数及(1.5)正好含m个条件,可以推出$p(t)$是唯一的.具体证明留作练习. □

(1.9)中的$p(t)$称为 **Lagrange-Hermite 插值公式**,$\lambda_1, \cdots, \lambda_s$称为**插值结点**,(1.5)称为**插值条件**.

插值公式$p(t)$的一般表达式如下:

$p(t)$

$$=\sum_{k=1}^{s}\left\{\left[\left[\sum_{j=0}^{m_k-1}\frac{1}{j!}\frac{d^j}{dt^j}\frac{f(t)}{\prod_{r=1,r\neq k}^{s}(t-\lambda_r)^{m_r}}\right|_{t=\lambda_k}(t-\lambda_k)^j\right]\prod_{r=1,r\neq k}^{s}(t-\lambda_r)^{m_r}\right\}$$

$$=\sum_{k=1}^{s}\left\{\left[\sum_{j=0}^{m_k-1}\frac{1}{j!}\varphi_k^{(j)}(\lambda_k)(t-\lambda_k)^j\right]\frac{\psi(t)}{(t-\lambda_k)^{m_k}}\right\}, \tag{1.12}$$

其中

$$\psi(t)\equiv(t-\lambda_1)^{m_1}\cdots(t-\lambda_s)^{m_s};\ \varphi_k(t)\equiv\frac{f(t)(t-\lambda_k)^{m_k}}{\psi(t)},$$

$$k=1,\cdots,s. \tag{1.13}$$

如果

$$m_1=\cdots=m_s=1,$$

那么 $s=m$,插值条件简化为

$$p(\lambda_k)=f(\lambda_k),\qquad k=1,\cdots,m. \tag{1.14}$$

插值公式 $p(t)$ 成为

$$p(t)=\sum_{k=1}^{m}f(\lambda_k)\frac{(t-\lambda_1)\cdots(t-\lambda_{k-1})(t-\lambda_{k+1})\cdots(t-\lambda_m)}{(\lambda_k-\lambda_1)\cdots(\lambda_k-\lambda_{k-1})(\lambda_k-\lambda_{k+1})\cdots(\lambda_k-\lambda_m)}. \tag{1.15}$$

这就是熟知的 **Lagrange 插值公式.**

插值多项式 $p(t)$ 是唯一的,但可以有不同的表示形式. $p(t)$ 的又一种表示基于函数 f 的均差,它无须区分插值结点的相异和相重的情形.

9.1.4 定义 设 f 是一元实或复值函数,$\lambda_1,\cdots,\lambda_m$ 是属于 f 定义域的实或复数,暂且假定它们是相异的,但最终允许有相重值.

令

$$\Delta^0 f(\lambda_1) \equiv f(\lambda_1),$$

$$\Delta f(\lambda_1,\lambda_2) \equiv \Delta^1 f(\lambda_1,\lambda_2) \equiv \frac{f(\lambda_1)-f(\lambda_2)}{\lambda_1-\lambda_2}. \tag{1.16}$$

并且递归地,

$$\Delta^k f(\lambda_1,\cdots,\lambda_k,\lambda_{k+1}) \equiv \frac{\Delta^{k-1} f(\lambda_1,\cdots,\lambda_{k-1},\lambda_k)-\Delta^{k-1} f(\lambda_1,\cdots,\lambda_{k-1},\lambda_{k+1})}{\lambda_k-\lambda_{k+1}}$$

$$k=1,\cdots,m-1. \tag{1.17}$$

$\Delta^k f(\lambda_1,\cdots,\lambda_{k+1})$称为关于结点$\lambda_1,\cdots,\lambda_{k+1}$的$k$**阶均差**.

9.1.5 定理 次数$\le m-1$的多项式

$$\begin{aligned} p(t)= {} & f(\lambda_1)+\Delta f(\lambda_1,\lambda_2)(t-\lambda_1) \\ & +\Delta^2 f(\lambda_1,\lambda_2,\lambda_3)(t-\lambda_1)(t-\lambda_2)+\cdots \\ & +\Delta^{m-1} f(\lambda_1,\cdots,\lambda_m)(t-\lambda_1)\cdots(t-\lambda_{m-1}) \end{aligned} \tag{1.18}$$

满足插值条件(1.14).

证 显然,

$$p(\lambda_1)=f(\lambda_1).$$

而且

$$\begin{aligned} p(\lambda_2) &= f(\lambda_1)+\Delta f(\lambda_1,\lambda_2)(\lambda_2-\lambda_1) \\ &= f(\lambda_1)-(f(\lambda_1)-f(\lambda_2))=f(\lambda_2). \end{aligned}$$

利用数学归纳法(留作练习),容易推出成立(1.14). □

(1.18)称为 **Newton 均差插值公式**.

根据这个定理,表达式(1.18)和 Lagrange 插值公式(1.15)应是相同的多项式.特别,两个公式的t^{k-1}的系数必须相等,从而得出恒等式

$$\Delta^{m-1} f(\lambda_1,\cdots,\lambda_m)$$
$$=\sum_{k=1}^{m}\frac{f(\lambda_k)}{(\lambda_k-\lambda_1)\cdots(\lambda_k-\lambda_{k-1})(\lambda_k-\lambda_{k+1})\cdots(\lambda_k-\lambda_m)}. \tag{1.19}$$

由此得知:如果$\lambda_1,\cdots,\lambda_m$是相异的,$f$是连续函数,那么$f$的每个均差是关于其结点对称而且连续的函数.

为了将 Newton 公式(1.18)和 Lagrange-Hermite 公式(1.12)联系起来,必须考虑怎样定义某些结点相重时的均差.由于结点相异时均差对其结点是连续的,要保持连续性,有结点相重时均差的定义只能有一种方式,例如

$$\Delta^{k-1} f(\lambda_1,\lambda_1,\lambda_3,\cdots,\lambda_k)\equiv\lim_{\lambda_2\to\lambda_1}\Delta^{k-1} f(\lambda_1,\lambda_2,\lambda_3,\cdots,\lambda_k), \tag{1.20}$$

条件是右端的极限存在.

9.1.6 定理 设f是实或复值函数.如果f在单连通开集$D\subset\mathbb{C}$上是解析的,$\lambda_1,\cdots,\lambda_k\in D$,或者,如果$f$在凸集$D\subset\mathbb{C}$上$k-1$次连续可微,$\lambda_1,\cdots,\lambda_k$在$D$的相对内部而且是共线的(特别,它们全为实数时,$D$也可以是一线段),那么可以保持连续性地将均差的定义扩展至允许有相重的结点;特别,

$$\Delta^{k-1} f(\lambda,\cdots,\lambda)$$
$$\equiv\lim_{\lambda_1\to\lambda,\cdots,\lambda_k\to\lambda}\Delta^{k-1} f(\lambda_1,\cdots,\lambda_k)=\frac{1}{(k-1)!}f^{(k-1)}(\lambda), \tag{1.21}$$

其中λ与$\lambda_1,\cdots,\lambda_k$一起是共线的.

证 如果f在单连通开集D上解析,取

$$\Gamma\subset D$$

是严格包围点$\lambda,\lambda_1,\cdots,\lambda_k$的简单可求长曲线.根据 Cauchy 积分定理,当$\lambda_1,\cdots,\lambda_k$相异时,有

$$\frac{1}{2\pi i}\oint_{\Gamma}\frac{f(t)\mathrm{d}t}{(t-\lambda_1)\cdots(t-\lambda_k)}$$

$$=\sum_{i=1}^{k}\frac{f(\lambda_i)}{(\lambda_i-\lambda_1)\cdots(\lambda_i-\lambda_{i-1})(\lambda_i-\lambda_{i+1})\cdots(\lambda_i-\lambda_k)}. \quad (1.22)$$

再根据(1.19),有

$$\Delta^{k-1}f(\lambda_1,\cdots,\lambda_k)=\frac{1}{2\pi i}\oint_{\Gamma}\frac{f(t)\mathrm{d}t}{(t-\lambda_1)\cdots(t-\lambda_k)}. \quad (1.23)$$

注意到右边的积分是$\lambda_1,\cdots,\lambda_k\in D$的连续函数,而无论$\lambda_1,\cdots,\lambda_k$相异或有相重,因此可以用(1.23)将均差定义扩充至有相重的结点.特别

$$\Delta^{k-1}f(\lambda,\cdots,\lambda)=\frac{1}{2\pi i}\oint_{\Gamma}\frac{f(t)\mathrm{d}t}{(t-\lambda)^k}=\frac{1}{(k-1)!}f^{(k-1)}(\lambda).$$

现在设f在凸集D上$k-1$次连续可微,且$\lambda,\lambda_1,\cdots,\lambda_k$共线.考虑累次积分

$$\phi_{k-1}(\lambda_1,\cdots,\lambda_k)$$
$$\equiv\int_0^1\cdots\int_0^{\tau_{k-3}}\int_0^{\tau_{k-2}}f^{(k-1)}(\lambda_1+(\lambda_2-\lambda_1)\tau_1+\cdots$$
$$+(\lambda_k-\lambda_{k-1})\tau_{k-1})\mathrm{d}\tau_{k-1}\cdots\mathrm{d}\tau_1, \quad (1.24)$$

这一积分是有意义的,因为其被积函数的自变量是$\lambda_1,\cdots,\lambda_k$的凸组合.如果点$\lambda_1,\cdots,\lambda_k$相异,作一次积分,得恒等式

$$\phi_{k-1}(\lambda_1,\cdots,\lambda_k)=\frac{\phi_{k-2}(\lambda_1,\cdots,\lambda_{k-2},\lambda_k)-\phi_{k-2}(\lambda_1,\cdots,\lambda_{k-1})}{\lambda_k-\lambda_{k-1}}.$$

依次对$k-2,k-3,\cdots$递推,最后可得

$$\phi_1(\lambda_1,\lambda_2)=\frac{f(\lambda_1)-f(\lambda_2)}{\lambda_1-\lambda_2}.$$

将这些恒等式与均差定义**9.1.4**比较,表明$\lambda_1,\cdots,\lambda_k$相异时,

$$\phi_{k-1}(\lambda_1,\cdots,\lambda_k)=\Delta^{k-1}f(\lambda_1,\cdots,\lambda_k).$$

注意到(1.24)的积分是$\lambda_1,\cdots,\lambda_k\in D$的连续函数,可以用

$$\begin{aligned}\Delta^{k-1}f(\lambda_1,\cdots,\lambda_k)\\ \equiv\int_0^1\cdots\int_0^{\tau_{k-3}}\int_0^{\tau_{k-2}}f^{(k-1)}(\lambda_1+(\lambda_2-\lambda_1)\tau_1+\cdots\\ +(\lambda_k-\lambda_{k-1})\tau_{k-1})\mathrm{d}\tau_{k-1}\cdots\mathrm{d}\tau_1\end{aligned}\tag{1.25}$$

将均差定义扩充至有相重的结点.特别

$$\Delta^{k-1}f(\lambda,\cdots,\lambda)\equiv\int_0^1\cdots\int_0^{\tau_{k-3}}\int_0^{\tau_{k-2}}f^{(k-1)}(\lambda)d\tau_{k-1}\cdots d\tau_1$$

$$=\frac{1}{(k-1)!}f^{(k-1)}(\lambda).\qquad\square$$

9.1.7 定理 设f是实或复值函数.

如果f在单连通开集$D\subset\mathbb{C}$上是解析的,令$\mathscr{D}\equiv D$,由(1.23)定义均差.

或者,如果f在凸集$D\subset\mathbb{C}$上$m-1$次连续可微,$\mathscr{D}$是D的相对内部,由(1.25)定义均差.

在以上两种假设的任一情况下,对于

$$\lambda_1,\cdots,\lambda_m\in\mathscr{D},$$

将仍由Newton公式(1.18)定义的多项式$p(t)$改写成

$$\begin{aligned}p(t)=\alpha_{m-1}(\lambda_1,\cdots,\lambda_m)t^{m-1}+\cdots\\ +\alpha_1(\lambda_1,\cdots,\lambda_m)t+\alpha_0(\lambda_1,\cdots,\lambda_m),\end{aligned}\tag{1.26}$$

那么

(1) 所有均差和(1.26)的系数

$$\Delta^{k-1}f(\lambda_1,\cdots,\lambda_k),\ \alpha_{k-1}(\lambda_1,\cdots,\lambda_m),\quad k=1,\cdots,m$$

是$\lambda_1,\cdots,\lambda_m\in\mathscr{D}$的连续的和对称的函数.

(2) 对于选取的点$\lambda_1,\cdots,\lambda_m$,若允许其中有相重的,并且设

$$\lambda_1,\cdots,\lambda_s$$

是其相异值,λ_k的相重数为m_k,$k=1,\cdots,s$,$m_1+\cdots+m_s=m$,

则

$$p^{(j)}(\lambda_k)=f^{(j)}(\lambda_k),\ \ j=0,1,\cdots,m_k-1;\quad k=1,\cdots,s.$$

所以 $p(t)$ 是 Lagrange-Hermite 插值问题(1.5)的解.

证 由(1.19)已知,当 f 连续时,对于相异的 $\lambda_1,\cdots,\lambda_m$, f 的每个均差是关于其结点对称而且连续的.现在,根据(1.23)或(1.25)定义均差,**9.1.6** 表明,在有相重结点时,仍保持其连续性;显然,连续性又保证了保持其对称性.因此,从 Newton 公式(1.18)经展开和并项而得的(1.26)的系数也必定是连续和对称的.结论(1)得证.

基于以上结论,可以按任何方便的顺序重新排列 $\lambda_1,\cdots,\lambda_m$ 而不会改变多项式 $p(t)$.现在,将 $\lambda_1,\cdots,\lambda_m$ 中 m_1 个相重的 λ_1 集中排列在前面,利用(1.23)或(1.25),有表示式

$$\begin{aligned}p(t)=f(\lambda_1)+\frac{f'(\lambda_1)}{1!}(t-\lambda_1)+\frac{f''(\lambda_1)}{2!}(t-\lambda_1)^2+\cdots\\+\frac{f^{(m_1-1)}(\lambda_1)}{(m_1-1)!}(t-\lambda_1)^{m_1-1}+R(t)(t-\lambda_1)^{m_1},\end{aligned}$$

其中 $R(t)$ 是多项式.推出

$$p^{(j)}(\lambda_1)=f^{(j)}(\lambda_1),\ \ j=0,1,\cdots,m_1-1.$$

一般,如果将 $\lambda_1,\cdots,\lambda_m$ 中 m_k 个相重的 λ_k 集中排列在前面,可推出

$$p^{(j)}(\lambda_k)=f^{(j)}(\lambda_k),\ \ j=0,1,\cdots,m_k-1.$$

结论(2)得证. □

9.2 多项式矩阵函数

9.2.1 定义 存在几种关于**多项式**的概念及其相伴的**多项式矩阵函数**的概念.

设 $p(t)$ 是通常的**纯量值**多项式,

$$p(t) \equiv a_m t^m + a_{m-1} t^{m-1} + \cdots + a_1 t + a_0, \quad a_0, \cdots, a_m \in \mathbb{C}, \tag{2.1}$$

可以自然地定义相伴 $p(t)$ 的**矩阵函数** $p(A)$,

$$p(A) \equiv a_m A^m + a_{m-1} A^{m-1} + \cdots + a_1 A + a_0 I, \quad \forall A \in \mathbb{C}^{n\times n}. \tag{2.2}$$

另一种情形,设 $P(t)$ 是**矩阵值**多项式,

$$P(t) \equiv A_m t^m + A_{m-1} t^{m-1} + \cdots + A_1 t + A_0, \quad A_0, \cdots, A_m \in \mathbb{C}^{n\times n}. \tag{2.3}$$

可以自然地定义两种相伴 $P(t)$ 的**矩阵函数** $P_l(A)$ 和 $P_r(A)$,

$$P_l(A) \equiv A^m A_m + A^{m-1} A_{m-1} + \cdots + A A_1 + A_0, \quad \forall A \in \mathbb{C}^{n\times n} \tag{2.4}$$

$$P_r(A) \equiv A_m A^m + A_{m-1} A^{m-1} + \cdots + A_1 A + A_0, \quad \forall A \in \mathbb{C}^{n\times n} \tag{2.5}$$

显然,除非 A 与所有系数矩阵 $A_0, \cdots, A_m$ 可交换,否则 $P_l(t)$ 和 $P_r(t)$ 一般是不同的.

$p(A)$, $P_l(A)$ 和 $P_r(A)$ 的计算常常可以简化,对于多项式次数 m 比矩阵阶数 n 大的情形来说尤其如此.因为依 Cayley-Hamilton 定理 1.6.9,给定矩阵 $A \in \mathbb{C}^{n\times n}$ 大于 $n-1$ 次的幂总可以表示成低次幂的线性组合,因此当 $m > n-1$ 时,形如(2.2),(2.4)和(2.5)的多项式函数总可以修改成至多包含 $A, A^2, \cdots, A^{n-1}$ 的表达式,然而修改后的形式依赖于矩阵 A.

9.2.2 定理　设 $A \in \mathbb{C}^{n\times n}$, $p(t)$ 是(2.1)给定的多项式.并设

$$A = SJS^{-1},$$

其中

$$J=\begin{bmatrix} J_{n_1}(\lambda_1) & & 0 \\ & \ddots & \\ 0 & & J_{n_s}(\lambda_s) \end{bmatrix} \tag{2.6}$$

是 A 的 Jordan 标准形,$J_{n_i}(\lambda_i)$是由关于 A 的特征值 λ_i 的若干 Jordan 块组成的$n_i \times n_i$矩阵,$i=1,\cdots,s$,则

$$p(A)=S\begin{bmatrix} p(J_{n_1}(\lambda_1)) & & 0 \\ & \ddots & \\ 0 & & p(J_{n_s}(\lambda_s)) \end{bmatrix}S^{-1}. \tag{2.7}$$

而且,一般地说,对于$k\times k$ Jordan 块$J_k(\lambda)$,

$$p(J_k(\lambda))$$
$$=\begin{bmatrix} p(\lambda) & \frac{1}{1!}p'(\lambda) & \frac{1}{2!}p''(\lambda) & \cdots & \cdots & \frac{1}{(k-1)!}p^{(k-1)}(\lambda) \\ & p(\lambda) & \frac{1}{1!}p'(\lambda) & \ddots & & \frac{1}{(k-2)!}p^{(k-2)}(\lambda) \\ & & p(\lambda) & \ddots & \ddots & \vdots \\ & & & \ddots & \ddots & \frac{1}{2!}p''(\lambda) \\ & 0 & & & p(\lambda) & \frac{1}{1!}p'(\lambda) \\ & & & & & p(\lambda) \end{bmatrix}. \tag{2.8}$$

(2.7)中的每个对角块

$$p(J_{n_i}(\lambda_i))$$

是关于$\lambda=\lambda_i$的以形如(2.8)的矩阵为块组成的块对角矩阵,$i=1,\cdots,s$.

证 根据(2.2),以及定理的假设,显然成立(2.7).因此,主要是推证(2.8).

注意到

$$J_k(\lambda) = \lambda I + N \in \mathbb{C}^{k\times k},$$

其中 $N \in \mathbb{C}^{k\times k}$ 是 k 次幂零矩阵,

$$N = \begin{bmatrix} 0 & 1 & & 0 \\ & \ddots & \ddots & \\ & & \ddots & 1 \\ 0 & & & 0 \end{bmatrix}; \qquad N^i = 0,\ i \ge k$$

于是有

$$J_k(\lambda)^j = (\lambda I + N)^j = \sum_{i=0}^{j} \binom{j}{i} \lambda^{j-i} N^i,$$

由此推出

$$\begin{aligned} p(J_k(\lambda)) &= \sum_{j=0}^{m} a_j J_k(\lambda)^j = \sum_{j=0}^{m} \sum_{i=0}^{j} a_j \binom{j}{i} \lambda^{j-i} N^i \\ &= \sum_{i=0}^{m} \left\{ \sum_{j=i}^{m} \binom{j}{i} a_j \lambda^{j-i} \right\} N^i = \sum_{i=0}^{m} \frac{1}{i!} \left\{ \sum_{j=i}^{m} \frac{j!}{(j-i)!} a_j \lambda^{j-i} \right\} N^i \\ &= \sum_{i=0}^{m} \frac{1}{i!} p^{(i)}(\lambda) N^i = \sum_{i=0}^{\min\{m,k-1\}} \frac{1}{i!} p^{(i)}(\lambda) N^i. \end{aligned}$$

最后的表达式的矩阵形式即为(2.8). □

特别,如果 $A \in \mathbb{C}^{n\times n}$ 可对角化,

$$A = S\Lambda S^{-1}, \quad \Lambda = \operatorname{diag}(\lambda_1, \cdots, \lambda_n), \tag{2.9}$$

那么

$$\begin{aligned} p(A) = p(S\Lambda S^{-1}) &= Sp(\Lambda)S^{-1} \\ &= S \begin{bmatrix} p(\lambda_1) & & 0 \\ & \ddots & \\ 0 & & p(\lambda_n) \end{bmatrix} S^{-1}. \end{aligned} \tag{2.10}$$

9.2.3 定理 设$\|\bullet\|$是$\mathbb{C}^{n\times n}$上的矩阵算子范数.

(1) 设$p(t)$是(2.1)给定的多项式,并设

$$p_{\text{abs}}(t)\equiv|a_m|t^m+|a_{m-1}|t^{m-1}+\cdots+|a_1|t+|a_0|, \quad (2.11)$$

则

$$\begin{aligned}\|p(A+E)-p(A)\| &\le p_{\text{abs}}(\|A\|+\|E\|)-p_{\text{abs}}(\|A\|)\\ &\le \|E\|p'_{\text{abs}}(\|A\|+\|E\|), \quad \forall A,E\in\mathbb{C}^{n\times n}, \quad (2.12)\end{aligned}$$

其中$p'_{\text{abs}}(t)$是$p_{\text{abs}}(t)$关于t的导数.

(2) 设$P_l(A)$和$P_r(A)$分别是(2.4)和(2.5)给定的矩阵系数的多项式,并设

$$P_{\text{abs}}(t)\equiv\|A_m\|t^m+\|A_{m-1}\|t^{m-1}+\cdots+\|A_1\|t+\|A_0\|, \quad (2.13)$$

则

$$\begin{aligned}\left.\begin{array}{l}\|P_l(A+E)-P_l(A)\|\\ \|P_r(A+E)-P_r(A)\|\end{array}\right\} &\le P_{\text{abs}}(\|A\|+\|E\|)-P_{\text{abs}}(\|A\|)\\ &\le \|E\|P'_{\text{abs}}(\|A\|+\|E\|), \quad \forall A,E\in\mathbb{C}^{n\times n}, \quad (2.14)\end{aligned}$$

其中$P'_{\text{abs}}(t)$是$P_{\text{abs}}(t)$关于t的导数.

总的说,$p(t)$,$P_l(A)$和$P_r(A)$都是$\mathbb{C}^{n\times n}$上的连续函数,在$\mathbb{C}^{n\times n}$的每一个紧子集K上是 Lipschitz 连续的,也就是说,对于每个如此的K,存在正数L,使得

$$\|f(B)-f(A)\|\le L\|B-A\|, \quad \forall A,B\in K, \quad (2.15)$$

这里用f代表p,P_l或P_r.

证 三种情况共同的基本事实是二项式$(A+E)^k$的非交换的展式.

为了便于说明,以$k=3$的情形为例:

$$\begin{aligned}(A+E)^3=A^3&+\left(A^2E+AEA+EA^2\right)\\ &+\left(AE^2+EAE+E^2A\right)+E^3.\end{aligned}$$

由此,并利用范数的三角不等式及次可加性,有

$$\left\|(A+E)^3-A^3\right\| \le 3\|A\|^2\|E\|+3\|A\|\|E\|^2+\|E\|^3$$
$$=(\|A\|+\|E\|)^3-\|A\|^3.$$

经简单的归纳推证,得

$$\left\|(A+E)^k-A^k\right\| \le (\|A\|+\|E\|)^k-\|A\|^k, k=1,2,\cdots \quad (2.16)$$

因此

$$\|p(A+E)-p(A)\| \le \sum_{k=0}^{m}|a_k|\left[(\|A\|+\|E\|)^k-\|A\|^k\right]$$
$$=p_{\text{abs}}(\|A\|+\|E\|)-p_{\text{abs}}(\|A\|)$$
$$=\|E\|p'_{\text{abs}}(\xi)\le\|E\|p'_{\text{abs}}(\|A\|+\|E\|),$$

其中

$$\|A\|\le\xi\le\|A\|+\|E\|.$$

以上不等式蕴涵 $p(A)$ 的连续性.

关于在紧子集 $K\in\mathbb{C}^{n\times n}$ 上 Lipschitz 连续,取

$$A,B\in K,\quad E=B-A,$$

即得

$$\|p(B)-p(A)\|=\|p(A+(B-A))-p(A)\|$$
$$\le\|B-A\|p'_{\text{abs}}(\|A\|+\|B-A\|)\le L\|B-A\|,$$

其中 $L\equiv\max\{p'_{\text{abs}}(\|A\|+\|B-A\|): A,B\in K\}$.

(2)的证明相仿. □

9.2.4 定理 设 $p(t)$ 和 $r(t)$ 是复系数纯量值多项式, $A\in\mathbb{C}^{n\times n}$,则 $p(A)=r(A)$ 的充分必要条件是 $p(t)$ 和 $r(t)$ 在 A 的谱 $\lambda(A)$ 上有相同的值(见 **9.1.2**).

证 设(1.2)中的 $m(t)$ 是 A 的最小多项式, 令

$$g(t)=p(t)-r(t). \quad (2.17)$$

如果 $p(A)=r(A)$,那么 $g(A)=0$.从而, $g(t)$ 可被 $m(t)$ 整除,亦即存在多项式 $q(t)$,使得

$$g(t)=q(t)m(t).$$

由此,成立

$$p^{(i)}(\lambda_k)-r^{(i)}(\lambda_k)=g^{(i)}(\lambda_k)=0,\ i=0,1,\cdots,m_k-1;$$

$$k=1,\cdots,s. \quad (2.18)$$

这表明 $p(t)$ 和 $r(t)$ 在 $\lambda(A)$ 上有相同的值.

反之,若(2.18)成立,则 $g(t)$ 以 λ_k 为 m_k 重零点, $k=1,\cdots,s$.因此 $g(t)$ 必为(1.2)中的 $m(t)$ 整除.鉴于 $m(t)$ 是 A 的最小多项式,推出 $g(A)=0$,等价地, $p(A)=r(A)$. □

根据这一定理及 **9.1.3** 可以得出这样的结论:对于任何给定的 $A\in\mathbb{C}^{n\times n}$ 及次数 $\geq n$ 的复系数纯量值多项式 $p(t)$,必存在次数 $\leq n-1$ 的多项式 $r(t)$,使得

$$p(A)=r(A). \quad (2.19)$$

9.3 非多项式矩阵函数

9.3.1 引言 现在考虑对于一般的一元复或实变量 t 的纯量值连续性函数 $f(\cdot)$ 以及 $A\in\mathbb{C}^{n\times n}$,应当怎样定义矩阵值函数 $f(A)$.

似乎很自然要求 $f(A)$ 是 A 的连续函数;而且如同(2.10)那样,当 A 可对角化, $A=S\Lambda S^{-1}$, $\Lambda=\operatorname{diag}(\lambda_1,\cdots,\lambda_n)$ 时,

$$f(A)=S\begin{bmatrix} f(\lambda_1) & & 0 \\ & \ddots & \\ 0 & & f(\lambda_n) \end{bmatrix}S^{-1}. \quad (3.1)$$

对于 A 可对角化的情形,(3.1)是否给予了 $f(A)$ 的值一个确定的定义,或者说它是否依赖于 A 的特征值的排序或对角化的相似矩

阵 S 的选择呢?

假设 $t_1,\cdots,t_s$ 是 A 的特征值 $\lambda_1,\cdots,\lambda_n$ 中的不同值,并设 A 的任一对角化如下:

$$A = TMT^{-1}, M = \mathrm{diag}(\mu_1,\cdots,\mu_n),$$

则 $\mu_1,\cdots,\mu_n$ 只是 $\lambda_1,\cdots,\lambda_n$ 的一个不同排序;另一方面,可以利用 Lagrange 公式(1.15)构造一个次数 $\leq s-1$ 的多项式 $p(t)$,使得

$$p(t_i) = f(t_i), \quad i = 1,\cdots,s.$$

因此

$$\begin{aligned} f(A) &= S\mathrm{diag}(f(\lambda_1),\cdots,f(\lambda_n))S^{-1} \\ &= S\mathrm{diag}(p(\lambda_1),\cdots,p(\lambda_n))S^{-1} \\ &= Sp(\Lambda)S^{-1} = p(S\Lambda S^{-1}) = p(A) \\ &= p(TMT^{-1}) = Tp(M)T^{-1} \\ &= T\mathrm{diag}(p(\mu_1),\cdots,p(\mu_n))T^{-1} \\ &= T\mathrm{diag}(f(\mu_1),\cdots,f(\mu_n))T^{-1}. \end{aligned}$$

这一结果表明(3.1)给定的 $f(A)$ 的值与用于表达 A 的对角化形式无关.

因为可对角化矩阵在 $\mathbb{C}^{n\times n}$ 中是稠密的,也就是说,$\mathbb{C}^{n\times n}$ 中任一矩阵都可以利用可对角化矩阵序列来逼近,所以对于定义 $\mathbb{C}^{n\times n}$ 上的 $f(A)$,要保持连续性,至多有一种方式(或者说达到一种结果).只要 $f(t)$ 足够光滑,这样一种方式是确实存在的.

会令人惊奇的是:当 $n\geq 2$ 时, $f(A)$ 是 $A\in\mathbb{C}^{n\times n}$ 的连续函数这一条件蕴涵 $f(t)$ 必是可微的.

设 $f(t)$ 在 $\lambda\in\mathbb{C}$ 的一个邻域内有定义.考虑可对角化矩阵

$$A_\varepsilon \equiv \begin{bmatrix} \lambda & 1 \\ 0 & \lambda+\varepsilon \end{bmatrix}, \quad \varepsilon \in \mathbb{C}, \ \varepsilon \neq 0.$$

对此,有

$$A_\varepsilon = S_\varepsilon \Lambda_\varepsilon S_\varepsilon^{-1},$$

其中

$$\Lambda_\varepsilon = \begin{bmatrix} \lambda & 0 \\ 0 & \lambda+\varepsilon \end{bmatrix}.$$

而

$$S_\varepsilon = \begin{bmatrix} 1 & 1 \\ 0 & \varepsilon \end{bmatrix}, \quad S_\varepsilon^{-1} = \frac{1}{\varepsilon}\begin{bmatrix} \varepsilon & -1 \\ 0 & 1 \end{bmatrix},$$

因此,对于充分小的$\varepsilon \neq 0$,根据(3.1)的要求,

$$\begin{aligned} f(A_\varepsilon) &= S_\varepsilon \begin{bmatrix} f(\lambda) & 0 \\ 0 & f(\lambda+\varepsilon) \end{bmatrix} S_\varepsilon^{-1} \\ &= \begin{bmatrix} f(\lambda) & \dfrac{f(\lambda+\varepsilon)-f(\lambda)}{\varepsilon} \\ 0 & f(\lambda+\varepsilon) \end{bmatrix}. \end{aligned} \tag{3.2}$$

由于

$$A_\varepsilon \to J_2(\lambda) \equiv \begin{bmatrix} \lambda & 1 \\ 0 & \lambda \end{bmatrix}, \quad 当\varepsilon \to 0,$$

如果$f(A)$在$A = J_2(\lambda)$连续,那么$f(t)$必须在$t=\lambda$连续且可微,而且$f(J_2(\lambda))$的值必定是

$$f(J_2(\lambda)) = \lim_{\varepsilon\to 0} f(A_\varepsilon) = \begin{bmatrix} f(\lambda) & f'(\lambda) \\ 0 & f(\lambda) \end{bmatrix}. \tag{3.3}$$

一般,如果$f(A)$对一切$n \geq 2$及在任一$A \in \mathbb{C}^{n\times n}$处连续,那么$f(t)$必须是一个解析函数.

现在设 $f(t)$ 在 $\lambda\in\mathbb{C}$ 为中心具有正半径 r 的小圆盘内解析.考虑可对角化矩阵

$$A_\varepsilon\equiv\begin{bmatrix}\lambda & 1 & & 0\\ & \lambda+\varepsilon & \ddots & \\ & & \ddots & 1\\ 0 & & & \lambda+(k-1)\varepsilon\end{bmatrix}\in\mathbb{C}^{k\times k},\varepsilon\neq 0\,. \quad (3.4)$$

必存在非奇异矩阵 $S_\varepsilon\in\mathbb{C}^{k\times k}$,使得

$$\begin{aligned}A_\varepsilon&=S_\varepsilon\Lambda_\varepsilon S_\varepsilon^{-1}\\&=S_\varepsilon\operatorname{diag}(\lambda,\lambda+\varepsilon,\cdots,\lambda+(k-1)\varepsilon)S_\varepsilon^{-1}.\end{aligned}\quad (3.5)$$

注意:

$$A_\varepsilon\to J_k(\lambda)\equiv\begin{bmatrix}\lambda & 1 & & 0\\ & \lambda & \ddots & \\ & & \ddots & 1\\ 0 & & & \lambda\end{bmatrix}\in\mathbb{C}^{k\times k},\text{当}\,\varepsilon\to 0,\quad (3.6)$$

$J_k(\lambda)$ 是一个 Jordan 块,是不可对角化的.

对于充分小的 $\varepsilon\neq 0$,设 $p_{A_\varepsilon}(t)$ 和 $p_{J_k(\lambda)}(t)$ 是 Newton 插值多项式. $p_{A_\varepsilon}(t)$ 满足

$$p_{A_\varepsilon}(\lambda+(j-1)\varepsilon)=f(\lambda+(j-1)\varepsilon),\ \ j=1,\cdots,k\,.$$

$p_{J_k(\lambda)}(t)$ 满足

$$p_{J_k(\lambda)}^{(j)}(\lambda)=f^{(j)}(\lambda),\ \ j=0,1,\cdots,k\,.$$

于是,根据(3.1)的要求和(3.5),

$$\begin{aligned}f(A_\varepsilon)&=S_\varepsilon(f(\lambda),f(\lambda+\varepsilon),\cdots,f(\lambda+(k-1)\varepsilon))S_\varepsilon^{-1}\\&=S_\varepsilon(p_{A_\varepsilon}(\lambda),p_{A_\varepsilon}(\lambda+\varepsilon),\cdots,p_{A_\varepsilon}(\lambda+(k-1)\varepsilon))S_\varepsilon^{-1}\\&=p_{A_\varepsilon}(A_\varepsilon).\end{aligned}\quad (3.7)$$

因为

$$p_{A_\varepsilon}(A_\varepsilon)=\{p_{A_\varepsilon}(A_\varepsilon)-p_{A_\varepsilon}(J_k(\lambda))\}$$

$$+\{p_{A_\varepsilon}(J_k(\lambda))-p_{J_k(\lambda)}(J_k(\lambda))\}+p_{J_k(\lambda)}(J_k(\lambda)).$$

根据 **9.2.3** 和(3.6),

$$p_{A_\varepsilon}(A_\varepsilon)-p_{A_\varepsilon}(J_k(\lambda))\to 0,\quad 当\,\varepsilon\to 0,$$

又根据 **9.1.7** 的(1),

$$p_{A_\varepsilon}(J_k(\lambda))\to p_{J_k(\lambda)}(J_k(\lambda)),\quad 当\,\varepsilon\to 0,$$

所以

$$p_{A_\varepsilon}(A_\varepsilon)\to p_{J_k(\lambda)}(J_k(\lambda)),\quad 当\,\varepsilon\to 0. \tag{3.8}$$

这样,从(3.7)和(3.8)得知,$f(A)$在 $A=J_k(\lambda)$连续的必要条件是

$$f(J_k(\lambda))=p_{J_k(\lambda)}(J_k(\lambda))$$

$$=\begin{bmatrix} f(\lambda) & f'(\lambda) & \frac{1}{2!}f''(\lambda) & \cdots & \frac{1}{(k-1)!}f^{(k-1)}(\lambda) \\ & f(\lambda) & f'(\lambda) & \ddots & \vdots \\ & & \ddots & \ddots & \frac{1}{2!}f''(\lambda) \\ & 0 & & \ddots & f'(\lambda) \\ & & & & f(\lambda) \end{bmatrix}. \tag{3.9}$$

经过以上概略的讨论,将条件(3.9)与(2.7)及(2.8)比较,可以较为自然地引出如下关于非多项式矩阵函数 $f(A)$ 的一般定义.

9.3.2 定义 设 $A\in\mathbb{C}^{n\times n}$,其谱为

$$\lambda(A)=\{\lambda_1,\cdots,\lambda_s:\lambda_i\neq\lambda_j,i\neq j\},$$

而且,A 的最小多项式为

$$m(t)=(t-\lambda_1)^{m_1}\cdots(t-\lambda_s)^{m_s}, \tag{3.10}$$

其中 $m_k\geq 1$ 是 A 的特征值 λ_k 的指数,$k=1,\cdots s$.并设

$$A = SJS^{-1}, \quad J \equiv \begin{bmatrix} J_{n_1}(\lambda_{s_1}) & & 0 \\ & \ddots & \\ 0 & & J_{n_r}(\lambda_{s_r}) \end{bmatrix}, \tag{3.11}$$

J 是 A 的 Jordan 标准形,其中

$$s \le r \le n;\ 1 \le s_i \le s,\ i = 1, \cdots, r;\ n_1 + \cdots + n_r = n.$$

每个对角块 $J_k(\lambda)$ 是特征值 λ 的 $k \times k$ Jordan 块, $k = n_1, \cdots, n_r$. 设 f 是 $D \subset \mathbb{C}$ 上的实或复值函数, $\lambda(A) \subset D$, 而且 f 是 $\lambda(A)$ 上的函数,也就是说(见 **9.1.2**),每个具有指数 $m_k > 1$ 的 λ_k 是 D 的内点, f 在 λ_k 是 $m_k - 1$ 次可微的.令

$$f(A) \equiv Sf(J)S^{-1} \equiv S \begin{bmatrix} f(J_{n_1}(\lambda_{s_1})) & & 0 \\ & \ddots & \\ 0 & & f(J_{n_r}(\lambda_{s_r})) \end{bmatrix} S^{-1}, \tag{3.12}$$

其中每个对角块定义为

$$f(J_k(\lambda)) \equiv \begin{bmatrix} f(\lambda) & f'(\lambda) & \frac{1}{2!}f''(\lambda) & \cdots & \frac{1}{(k-1)!}f^{(k-1)}(\lambda) \\ & f(\lambda) & f'(\lambda) & \ddots & \vdots \\ & & \ddots & \ddots & \frac{1}{2!}f''(\lambda) \\ & 0 & & \ddots & f'(\lambda) \\ & & & & f(\lambda) \end{bmatrix}. \tag{3.13}$$

如此矩阵值函数 $f(A)$ 称为相伴**纯量值干函数**(stem function) f 的**准素矩阵函数**(primary matrix function).

从定义看出,准素矩阵函数 $f(A)$,对于取定的 $A \in \mathbb{C}^{n \times n}$,仅涉及干函数 f 在 $\lambda(A)$ 或在 $\lambda(A)$ 附近的局部性质.

9.2.2 和 **9.2.4** 表明,当 $f(t)$ 是多项式时,对于任何 $A \in \mathbb{C}^{n\times n}$,准素矩阵函数 $f(A)$ 和(2.2)的多项式矩阵函数有相同的值;当 $A \in \mathbb{C}^{n\times n}$ 可对角化时,无论 $f(t)$ 是否多项式,准素矩阵函数 $f(A)$ 和(3.1)有相同的值.

介乎 $f(t)$ 是多项式和 A 可对角化两个极端之间成立同样结论的一种重要情形,概括成如下定理.

9.3.3 定理 设 f 是实或复值解析函数,具有幂级数表示

$$f(t) = a_0 + a_1 t + a_2 t^2 + \cdots \tag{3.14}$$

其收敛半径 $r > 0$.如果 $A \in \mathbb{C}^{n\times n}$, $\rho(A) < r$,那么矩阵幂级数

$$f(A) \equiv a_0 I + a_1 A + a_2 A^2 + \cdots \tag{3.15}$$

按 $\mathbb{C}^{n\times n}$ 上的任一种范数均收敛,而且其和等于相伴于函数 f 的准素矩阵函数 $f(A)$.

证 依 **2.1.22**,存在 $\mathbb{C}^{n\times n}$ 上的矩阵范数 $\|\cdot\|$ 使得 $\|A\| < r$.由此推出无穷级数(3.15)关于 $\mathbb{C}^{n\times n}$ 上的任一种范数均收敛.特别, $f(A)$ 的每个元素的数项级数是绝对收敛的.如果 A 的 Jordan 标准形由(3.11)给出,将 $A = SJS^{-1}$ 代入(3.14),得知 $f(A)$ 满足(3.12).最后,仿 **9.2.2** 中的计算,表明

$$f(J_k(\lambda)) = \sum_{i=0}^{\infty} \frac{1}{i!} f^{(i)}(\lambda) N^i = \sum_{i=0}^{k-1} \frac{1}{i!} f^{(i)}(\lambda) N^i$$

其中应用了 $N \equiv J_k(0), N^k = 0$;这正好是(3.13). □

9.3.4 定理 设 $A \in \mathbb{C}^{n\times n}$,其最小多项式为

$$m(t) = (t-\lambda_1)^{m_1} \cdots (t-\lambda_s)^{m_s},$$

$\lambda_1, \cdots, \lambda_s$ 是相异的, $m_k \ge 1$, $k = 1, \cdots, s$.设 f 和 g 是定义域包含点 $\lambda_1, \cdots, \lambda_s$ 的实或复值函数.每个具有 $m_k > 1$ 的 λ_k 是 f 和 g 的定义域的内点,而且 f 和 g 在 λ_k 处 $m_k - 1$ 次可微.设 $f(A)$ 和 $g(A)$ 分别

是相伴于函数 f 和 g 的准素矩阵函数.则

(1) 存在次数 $\le n-1$ 的多项式 $p(t)$,使得

$$f(A)=p(A),$$

$p(t)$ 可以取在 $m(t)=0$ 的根处插入 f 及其有关导数值的任何多项式.

(2) 准素矩阵函数 $f(A)$ 是有定义的,也就是说,$f(A)$ 的值与表示 A 所用的具体的 Jordan 标准形无关.

(3) 对于任何非奇异矩阵 $T\in\mathbb{C}^{n\times n}$,成立

$$f\left(TAT^{-1}\right)=Tf(A)T^{-1}.$$

(4) $f(A)$ 能与可同 A 交换的任何矩阵交换.

(5) $g(A)=f(A)$ 的充分必要条件是

$$g^{(i)}(\lambda_k)=f^{(i)}(\lambda_k),\ i=0,1,\cdots,m_k-1;\quad k=1,\cdots,s.$$

(6) 若 $A=A_1\oplus\cdots\oplus A_k$,则

$$f(A)=f(A_1)\oplus\cdots\oplus f(A_k).$$

(7) 若 A 有 Jordan 标准形

$$J_{n_1}(\lambda_{s_1})\oplus\cdots\oplus J_{n_r}(\lambda_{s_r}),$$

其中

$$s\le r\le n;\quad 1\le s_i\le s,\ i=1,\cdots,r;\quad n_1+\cdots+n_r=n,$$

则 $f(A)$ 的 Jordan 标准形和

$$f\left(J_{n_1}(\lambda_{s_1})\right)\oplus\cdots\oplus f\left(J_{n_r}(\lambda_{s_r})\right)$$

的 Jordan 标准形相同.特别,$f(A)$ 的特征值为 $f(\lambda_{s_1})\cdots,f(\lambda_{s_r})$,代数重数分别为 $n_1,\cdots,n_r$.

证 考虑 Lagrange-Hermite 公式(1.12)给出的多项式 $p(t)$,则有

$$p^{(i)}(\lambda_k)=f^{(i)}(\lambda_k),\ i=0,1,\cdots,m_k-1;\quad k=1,\cdots,s.$$

比较(2.7),(2.8)和(3.12),(3.13),表明 $f(A)=p(A)$.

为了证明(2),设

$$A=SJS^{-1}=T\hat{J}T^{-1},$$

其中 J 和 $\hat{J}$ 是 A 的 Jordan 标准形.因此,存在(块)置换矩阵 P,使得 $\hat{J}=PJP^{\mathrm{T}}$.根据(3.12)和(3.13),

$$f(\hat{J})=Pf(J)P^{\mathrm{T}}.$$

又根据(1),存在多项式 $p(t)$,使得 $p(J)=f(J)$.于是

$$\begin{aligned}Tf(\hat{J})T^{-1}&=TPf(J)P^{\mathrm{T}}T^{-1}=(TP)p(J)(TP)^{-1}\\&=p\left((TP)J(TP)^{-1}\right)=p\left(TPJP^{\mathrm{T}}T^{-1}\right)\\&=p\left(T\hat{J}T^{-1}\right)=p(A)=p\left(SJS^{-1}\right)\\&=Sp(J)S^{-1}=Sf(J)S^{-1}.\end{aligned}$$

对于结论(3),设 $A=SJS^{-1}$,J 是 A 的 Jordan 标准形,则有

$$TAT^{-1}=(TS)A(TS)^{-1},$$

从而

$$f\left(TAT^{-1}\right)\equiv(TS)f(J)(TS)^{-1}=T\left(Sf(J)S^{-1}\right)T^{-1}\equiv Tf(A)T^{-1}.$$

结论(4)可从(1)推出,因为与 A 可交换的矩阵必与 A 的任何多项式可交换.

结论(5)直接得自定义 **9.3.2**.

结论(6)和(7)也直接得自定义 **9.3.2**,只要把约化直和为 Jordan 标准形的相似矩阵,选作直接约化被加项为 Jordan 标准形的各个相似矩阵的直和. □

准素矩阵函数 $f(A)$ 是 A 的多项式这一事实很重要,然而同样重要的是对这种依赖于 A 的多项式的理解.并无必要存在单一的多项式 $p(t)$,使得对于 $f(X)$ 有定义的所有 $X\in\mathbb{C}^{n\times n}$ 成立 $p(X)=f(X)$.

例 考虑 $f(t)=\cos t$.设 $A\in\mathbb{C}^{n\times n}$ 的最小多项式为

$$m(t)=(t-1)^2(t-2),$$

依 **9.3.4** 的(1),任何多项式 $p(t)$,只要满足

$$p(1)=f(1)=\cos 1,\quad p'(1)=f'(1)=-\sin 1$$

和

$$p(2)=f(2)=\cos 2,$$

就有 $\cos A=p(A)$.可以利用 Lagrange-Hermite 公式(1.12),取

$$p(t)=(t-2)\left[\left.\frac{\cos t}{t-2}\right|_{t=1}+\frac{\mathrm{d}}{\mathrm{d}t}\left(\frac{\cos t}{t-2}\right)\right|_{t=1}(t-1)\right]$$

$$+(t-1)^2\left(\left.\frac{\cos t}{(t-1)^2}\right|_{t=2}\right)$$

$$=(t-2)[-\cos 1+(\sin 1-\cos 1)(t-1)]+(t-1)^2\cos 2.$$

这是 2 次多项式,对于最小多项式为 $(t-1)^2(t-2)$ 的阶 $n\geq 3$ 的任何矩阵 $A\in\mathbb{C}^{n\times n}$,成立 $p(A)=\cos A$.显然,对于最小多项式为 $(t-1)^2(t-2)$ 的因子的 $A\in\mathbb{C}^{n\times n}$,也成立 $p(A)=\cos A$.

9.3.5 推论 设 $A\in\mathbb{C}^{n\times n}$,$f$ 和 g 满足 **9.3.4** 的假设,则

(1) 如果 $f(t)\equiv c$,$c\in\mathbb{C}$,那么 $f(A)=cI$.

(2) 如果 $f(t)=t$,那么 $f(A)=A$.

(3) 如果 $h(t)\equiv f(t)+g(t)$,那么

$$h(A)=f(A)+g(A).$$

(4) 如果 $h(t)\equiv f(t)g(t)$,那么

$$h(A)=f(A)g(A)=g(A)f(A).$$

(5) 如果

$$g(\lambda)\neq 0,\forall\lambda\in\lambda(A);\quad h(t)\equiv f(t)/g(t),$$

那么 $g(A)$ 是非奇异的,而且

$$h(A)=f(A)[g(A)]^{-1}=[g(A)]^{-1}f(A).$$

证 (1),(2)和(3)直接由定义 **9.3.2** 推出.

对于(4),依 **9.3.4** 的(1),构造多项式 $p_1(t)$ 和 $p_2(t)$,使得

$$f(A)=p_1(A),\qquad g(A)=p_2(A).$$

令

$$p(t) \equiv p_1(t)p_2(t),$$

9.3.4 的(5)保证,在 A 的特征多项式的零点处,$p_1(t)$和 $p_2(t)$分别插入 $f(t)$和 $g(t)$的值及其导数值.利用乘积求导法则计算得知,$h(t)$在 A 的特征值λ_k 处存在直至m_k-1阶导数,而且

$$h^{(i)}(\lambda_k)=p^{(i)}(\lambda_k),\ i=0,1,\cdots,m_k-1;\quad k=1,\cdots,s,$$

因此,$h(A)$是有定义的,并依 **9.3.4** 的(5),

$$h(A)=p(A)=p_1(A)p_2(A)=f(A)g(A),$$

从 $f(A)$和 $g(A)$都是 A 的多项式推出两者是可交换的.

现在证明(5).令

$$\gamma(t)\equiv 1/g(t),$$

注意到在 A 的特征值处,$\gamma(t)$ 和 $g(t)$具有同样的可导性,故准素矩阵函数 $\gamma(A)$是有定义的.因为$1=g(t)\gamma(t)$,从(1)有

$$I=g(A)\gamma(A),$$

所以$\gamma(A)=[g(A)]^{-1}$.于是,从(4)推出(5). □

例 设 $A\in\mathbb{C}^{n\times n}$ 非奇异,$f(t)=1/t$.

由于 f 的定义域为 $D=\mathbb{C}-\{0\}$, f 在 D 上解析, A 的每个特征值是 D 的内点,故 $f(A)$ 可以作为准素矩阵函数求值.依 **9.3.5** 的(5),

$$f(A)=A^{-1},$$

亦即通常的矩阵之逆.因为

$$f^{(i)}(t)=(-1)^i\frac{i!}{t^{i+1}},\quad i=0,1,2,\cdots$$

所以(3.13)确定的非奇异 Jordan 块 $J_k(\lambda)$之逆是第1行为

$$\frac{1}{\lambda},\frac{-1}{\lambda^2},\frac{1}{\lambda^3},\cdots,\frac{(-1)^{k+1}}{\lambda^k}$$

的上三角 Toeplitz 矩阵 $f(J_k(\lambda))$.

9.3.6 推论　设 $A\in\mathbb{C}^{n\times n}$ 和 g 满足 **9.3.4** 的假设;设 f 是 $D'\subset\mathbb{C}$上的实或复值函数,

$$g(\lambda_1),\cdots,g(\lambda_s)\in D',$$

而且对于每个 $m_k>1$, $g(\lambda_k)$ 是 D' 的内点, f 在点 $g(\lambda_k)$ 处是 m_k-1 次可微的.记 $h(t)\equiv f(g(t))$,则相伴于函数 h 的准素矩阵函数 $h(A)$是有定义的,而且

$$h(A)=f(g(A)),\tag{3.16}$$

这里等式右端是两个准素矩阵函数

$$A\to g(A)\quad 和\quad g(A)\to f(g(A))$$

的复合.

证　设 A 的 Jordan 标准形 J 由(3.11)给出,于是

$$g(A)=Sg(J)S^{-1}=S\begin{bmatrix} g(J_{n_1}(\lambda_{s_1})) & & 0 \\ 0 & \ddots & \\ & & g(J_{n_r}(\lambda_{s_r})) \end{bmatrix}S^{-1}.$$

依 **9.3.4** 的(7), $g(A)$的 Jordan 标准形是

$$g(J_{n_1}(\lambda_{s_1})),\cdots,g(J_{n_r}(\lambda_{s_r}))$$

的 Jordan 标准形的直和.

但是, $g(J_{n_i}(\lambda_{s_i}))$的 Jordan 标准形中的 Jordan 块至多为 n_i 阶,不会大于 A 的最小多项式中 λ_{s_i} 的指数, $i=1,\cdots,r$.

因此,关于 $g(t)$和 $f(t)$的定义域与可微性的假设,充分保证按 **9.3.2** 而言的准素矩阵函数 $g(A)$和 $f(g(A))$是有定义的.

类似地,反复应用链式法则表明, h 所满足的定义域和可微性的条件,对于按 **9.3.2** 定义准素矩阵函数来说是必要的.利用 Lagrange-Hermite 公式(1.12)构造多项式 $p_1(t)$和 $p_2(t)$,使得

$$p_1^{(i)}(g(\lambda_k))=f^{(i)}(g(\lambda_k)),\quad p_2^{(i)}(\lambda_k)=g^{(i)}(\lambda_k);$$

$$i=0,1,\cdots,m_k-1;\quad k=1,\cdots,s.$$

从而,依 **9.3.4** 的(5),有

$$p_1(g(A)) = f(g(A)), \quad p_2(A) = g(A),$$

推出(3.17)

$$p_1(p_2(A)) = p_1(g(A)) = f(g(A)). \tag{3.17}$$

现在考虑多项式 $p_3(t) \equiv p_1(p_2(t))$.由于是多项式的复合,成立

$$p_3(A) = p_1(p_2(A)), \tag{3.18}$$

反复利用链式法则,表明

$$p_3^{(i)}(\lambda_k) = h^{(i)}(\lambda_k), \ i = 0,1,\cdots,m_k - 1; \quad k = 1,\cdots,s.$$

故仍依 **9.3.4** 的(5),有 $h(A) = p_3(A)$.由此及(3.17)和(3.18)得(3.16). □

9.3.7 推论 设 $A \in \mathbb{C}^{n\times n}$,其最小多项式为

$$m(t) = (t-\lambda_1)^{m_1} \cdots (t-\lambda_s)^{m_s},$$

$\lambda_1,\cdots,\lambda_s$ 是相异的,$m_k \ge 1, k = 1,\cdots,s$.设 f 是 $D \subset \mathbb{C}$ 上的实或复值函数,点 $t_1,\cdots,t_s \in D$ 满足

$$f(t_k) = \lambda_k, \quad k = 1,\cdots,s.$$

对于每个具有 $m_k > 1$ 的 t_k,或者 f 是在 t_k 的一个(二维复)邻域上的复变量解析函数;或者 f 是在 t_k 的一个(一维实)邻域上的实变量实值函数,而且在 t_k 处 $m_k - 1$ 次可微.并且设对于每个 $m_k > 1$ 有 $f'(t_k) \ne 0$,则存在 $X_0 \in \mathbb{C}^{n\times n}$,使得

$$f(X_0) = A.$$

而且,存在实或复值十函数 g,使得

$$X_0 = g(A)$$

是一个准素矩阵函数,从而 X_0 是 A 的一个多项式.

证 如果 $m_k = 1, k = 1,\cdots,s$,则无须再证.

现设 k 是一个指标,$m_k > 1$.在定理的假设下,逆函数定理保证存在 λ_k 的一个开邻域 $\mathscr{N}_k$ 及定义在 $\mathscr{N}_k$ 上的一个函数 g_k,使得

$f(g_k(s)) \equiv s$,而且 g_k 在 $s=\lambda_k$ 处 m_k-1 次可微.对于 $m_k=1$ 的 k,令 $\mathscr{N}_k \equiv \{\lambda_k\}$.不失一般性,假定 $\mathscr{N}_1,\cdots,\mathscr{N}_s$ 是互不相交的.令

$$D' \equiv \bigcup_{k=1}^{s} \mathscr{N}_k . \tag{3.19}$$

对于 $s \in D'$,定义 g 为

$$g(s) \equiv g_k(s), \quad \forall s \in \mathscr{N}_k; \quad k=1,\cdots,s, \tag{3.20}$$

则 $f(g(s)) \equiv s$, $\forall s \in D'$.故准素矩阵函数 $g(A)$ 是有定义的,而且 **9.3.6** 保证 $f(g(A)) \equiv A$. □

例 (1) 设 $A \in \mathbb{C}^{n\times n}$ 非奇异, $f(t)=t^2$,定义域 $D=\mathbb{C}$.

因为 $f'(t) \neq 0$, $\forall t \neq 0$,依 **9.3.7**,存在 $X_0 \in \mathbb{C}^{n\times n}$,使得

$$X_0^2 = A,$$

X_0 可以取作 A 的准素矩阵函数, $X_0 = g(A)$,而且

$$g(s) = \sqrt{s}, \quad \forall s \in D',$$

这里 D' 如同(3.19).

注意到 $g(A)$ 是 A 的多项式,故可得结论:每个非奇异矩阵 $A \in \mathbb{C}^{n\times n}$ 具有 A 的多项式形式的平方根($X^2=A$ 的解).从而,如果非奇异矩阵 A 是对称的或正规的,那么由 $g(A)$ 给出的平方根也是对称的或正规的,因为 A 的任何多项式保持 A 的这些性质.然而,并不是非奇异矩阵的任何平方根全都是准素矩阵函数.例如,矩阵 $\begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix}$ 是单位矩阵 $I=\begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$ 的平方根,但它不是 I 的多项式,因此不可能是 I 的准素矩阵函数.

(2) 设 $A \in \mathbb{C}^{n\times n}$ 非奇异, $f(t)=e^t$.

因为 $f'(t) \neq 0$, $\forall t \in \mathbb{C}$.依 **9.3.7**,存在 $X_0 \in \mathbb{C}^{n\times n}$,使得

$$e^{X_0} = A,$$

X_0 可以取作 A 的准素矩阵函数, $X_0 = g(A)$,而且

$$g(s)=\log s,\ \forall s\in D',$$

这里D'如同(3.19).

故可得结论:每个非奇异矩阵$A\in\mathbb{C}^{n\times n}$具有$A$的多项式形式的对数($e^X=A$的解).因为$f(t)=e^t$是整解析函数,其相伴的准素矩阵函数可以根据**9.3.2**或者利用e^t的幂级数来求值.

(3) 设$f(t)=t^k$,k是非零整数.并设$g(t)=e^t$.

令$D'=\mathbb{C}-\{0\}$,则g在$D=\mathbb{C}$上是解析的,$g(D)=D'$,而且$h(t)\equiv f(g(t))=e^{kt}$.依**9.3.6**,

$$e^{kA}=h(A)=f(g(A))=\left(e^A\right)^k,\forall A\in\mathbb{C}^{n\times n};$$
$$k=\pm1,\pm2,\cdots$$

取$k=-1$,得$e^{-A}=\left(e^A\right)^{-1}$;

(4) $\cos t\equiv\left(e^{it}+e^{-it}\right)/2$和$\sin t\equiv\left(e^{it}-e^{-it}\right)/2$是整解析函数,而且

$$\cos^2 t+\sin^2 t=1,\ \forall t\in\mathbb{C},$$

对于$\cos A$和$\sin A$可以利用它们的幂级数或作为准素矩阵函数求值.依**9.3.5**,

$$\cos^2 A+\sin^2 A=I,\ \forall A\in\mathbb{C}^{n\times n}.$$

(5) 以上例子从函数恒等式推出矩阵恒等式,然而反过来也是可能的.

设f和g是实或复变量t的实或复值函数,在它们定义域的内点$t=\lambda$处k次可微.并设$h(t)\equiv f(t)g(t)$,$A\equiv J_{k+1}(\lambda)$是$k+1$阶Jordan块.

依 **9.3.5** 的(4),$h(A)=f(A)g(A)$;$h(A)$,$f(A)$和$g(A)$均按(3.13)求值.$h(A)$的$(1,k+1)$-元素是$h^{(k)}(\lambda)/k!$,应恒等于乘积$f(A)g(A)$的$(1,k+1)$-元素,得到乘积求导的Leibnitz法则

$$\frac{\mathrm{d}^k}{\mathrm{d}t^k}\left[f(t)g(t)\right]=\sum_{m=0}^{k}\binom{k}{m}f^{(m)}(t)g^{(k-m)}(t),\quad k=1,2,\cdots \tag{3.21}$$

9.3.8 定理 设 $A,B\in\mathbb{C}^{n\times n}$ 是可交换的,则

$$\mathrm{e}^{A+B}=\mathrm{e}^{A}\mathrm{e}^{B}=\mathrm{e}^{B}\mathrm{e}^{A}. \tag{3.22}$$

特别,$\forall A\in\mathbb{C}^{n\times n}$ 成立

$$\mathrm{e}^{-A}=\left(\mathrm{e}^{A}\right)^{-1}.$$

而且

$$\mathrm{e}^{mA}=\left(\mathrm{e}^{A}\right)^{m},\quad m=\pm1,\pm2,\cdots \tag{3.23}$$

证 利用 e^{t} 的幂级数进行计算,

$$\begin{aligned}\mathrm{e}^{A+B}&=\sum_{k=0}^{\infty}\frac{1}{k!}(A+B)^k=\sum_{k=0}^{\infty}\frac{1}{k!}\sum_{j=0}^{k}\binom{k}{j}A^jB^{k-j}\\&=\sum_{k=0}^{\infty}\sum_{j=0}^{\infty}\frac{1}{j!(k-j)!}A^jB^{k-j}=\sum_{q=0}^{\infty}\sum_{p=0}^{\infty}\frac{1}{q!}\frac{1}{p!}A^qB^p\\&=\left(\sum_{q=0}^{\infty}\frac{1}{q!}A^q\right)\left(\sum_{p=0}^{\infty}\frac{1}{p!}B^p\right)=e^Ae^B.\end{aligned}$$

由此进一步推出

$$\mathrm{e}^{A}\mathrm{e}^{B}=\mathrm{e}^{A+B}=\mathrm{e}^{B+A}=\mathrm{e}^{B}\mathrm{e}^{A}.$$

特别,取 $B=-A$,有

$$I=\mathrm{e}^{0}=\mathrm{e}^{A-A}=\mathrm{e}^{A}\mathrm{e}^{-A},$$

取 $B=A$,有

$$\mathrm{e}^{2A}=\mathrm{e}^{A+A}=\mathrm{e}^{A}\mathrm{e}^{A}=\left(\mathrm{e}^{A}\right)^{2},$$

直接归纳得(3.23). □

9.3.9 定义 设 n 是正整数,$D\subset\mathbb{C}$.

$\mathscr{A}_n(D)$ 是 $D\to\mathbb{C}$ 的**函数集合**,定义如下:

$$\mathscr{A}_n(D) \equiv \begin{cases} C(D), & \text{当 } n=1, \\ C^{\infty}(D), & \text{当 } D \text{ 是单连通开集}, n \geq 2, \\ C^{n-1}(D), & \text{当 } D=(a,b) \subset \mathbb{R}, n \geq 2. \end{cases} \tag{3.24}$$

这里$C^k(D)$是通常的关于连续函数集合的记号,$f \in C^k(D)$表示$f: D \to \mathbb{C}$是k次连续可微的函数.特别,$C(D) \equiv C^0(D)$,即D上连续函数全体;$C^{\infty}(D)$是D上无穷可微函数,即解析函数的全体.

$\mathscr{D}_n(D)$是谱包含在D中的矩阵集合,定义如下:

$$\mathscr{D}_n(D) \equiv \{A \in \mathbb{C}^{n \times n} : \lambda(A) \subset D\}. \tag{3.25}$$

显然,对于每个$f \in \mathscr{A}_n(D)$来说,相伴以f为干函数的准素矩阵函数$f(A)$对于所有$A \in \mathscr{D}_n(D)$是有意义的.

9.3.10 定理 设n, p, q是正整数,$D \subset \mathbb{C}$.当$n \geq 2$时,设D或者是$\mathbb{C}$的单连通开集或者是$\mathbb{R}$的开区间,则

(1) 对于每个$f \in \mathscr{A}_n(D)$,准素矩阵函数$f(A)$在$\mathscr{D}_n(D)$上连续.

(2) 若$f: \mathscr{D}_n(D) \to \mathbb{C}^{p \times q}$和$g: \mathscr{D}_n(D) \to \mathbb{C}^{p \times q}$是连续函数,则

$$f(A) = g(A), \quad \forall A \in \mathscr{D}_n(D)$$

的充分必要条件是

$$f(A) = g(A), \quad \forall \text{ 可对角化的 } A \in \mathscr{D}_n(D).$$

证 $n=1$的情形是显然的.以下假定$n \geq 2$.

设$p_A(t)$是$f(t)$关于以A的特征值为结点的Newton插值公式.注意到

$$f(A) = p_A(A), \quad \forall A \in \mathscr{D}_n(D).$$

仿照(3.8)的证明,成立

$$f(A_\varepsilon) = p_{A_\varepsilon}(A_\varepsilon) \to p_A(A) = f(A), \text{ 当 } A_\varepsilon \to A$$

结论(1)得证.

由于 f 和 g 连续以及可对角化矩阵在 $\mathscr{D}_n(D)$ 内稠密,直接推出结论(2);这里 $f(A)$ 和 $g(A)$ 无须为准素矩阵函数. □

9.3.11 定理 设 n 是正整数, f 是单连通开集 $D\subset\mathbf{C}$ 上的复值解析函数, $f(A)$ 是相伴于函数 f 的准素矩阵函数,则对于每个 $A\in\mathscr{D}_n(D)$,

$$f(A)=\frac{1}{2\pi\mathrm{i}}\oint_{\Gamma}f(t)(tI-A)^{-1}\,\mathrm{d}t, \tag{3.26}$$

其中 $\Gamma\subset D$ 是严格包围 A 的所有特征值的简单可求长曲线.

证 记

$$\Phi(A)=\frac{1}{2\pi\mathrm{i}}\oint_{\Gamma}f(t)(tI-A)^{-1}\,\mathrm{d}t,$$

组成 $\boldsymbol{\Phi}(A)$ 的 n^2 个按元素的线积分是有定义的,因为 A 没有特征值在 Γ 上, $(tI-A)^{-1}$ 是 $t\in\Gamma$ 的连续函数.而且, A 包含在开邻域 $\mathscr{N}(A)\subset\mathscr{D}_n(D)$ 内,用 $\overline{\mathscr{N}(A)}$ 表示 $\mathscr{N}(A)$ 的闭包(是紧集),每个 $X\in\overline{\mathscr{N}(A)}$ 的所有特征值必严格在 Γ 所围区域的内部.由于 Γ 是紧集,积集 $\Gamma\times\overline{\mathscr{N}(A)}$ 也是紧的.因此,连续函数 $(tI-X)^{-1}$ 在 $\Gamma\times\overline{\mathscr{N}(A)}$ 上是一致有界的.

于是,恒等式

$$(tI-A)^{-1}-(tI-X)^{-1}=(tI-X)^{-1}(A-X)(tI-A)^{-1},$$

表明当 $X\to A$ 时,在 $t\in\Gamma$ 内

$$(tI-X)^{-1}\to(tI-A)^{-1}$$

是一致的,由此推出 $\boldsymbol{\Phi}(A)$ 是 $\mathscr{D}_n(D)$ 上的连续函数.

如果 $A\in\mathscr{D}_n(D)$ 是可对角化的,

$$A=S\Lambda S^{-1},\ \ \Lambda=\mathrm{diag}(\lambda_1,\cdots,\lambda_n),$$

依 Cauchy 积分定理,

$$\Phi(A)=\frac{1}{2\pi\mathrm{i}}\oint_{\Gamma}f(t)(tI-A)^{-1}\mathrm{d}t=\frac{1}{2\pi\mathrm{i}}\oint_{\Gamma}f(t)\left(tI-S\Lambda S^{-1}\right)^{-1}\mathrm{d}t$$

$$=S\mathrm{diag}\left(\frac{1}{2\pi\mathrm{i}}\oint_{\Gamma}f(t)(t-\lambda_1)^{-1}\mathrm{d}t,\cdots,\frac{1}{2\pi\mathrm{i}}\oint_{\Gamma}f(t)(t-\lambda_n)^{-1}\mathrm{d}t\right)S^{-1}$$

$$=S\mathrm{diag}(f(\lambda_1),\cdots,f(\lambda_n))S^{-1}=f(A).$$

至此,应用 **9,3,10**,得

$$f(A)=\Phi(A),\ \ \forall A\in\mathscr{D}_n(D).$$ □

9.3.12 定理 设$\|\bullet\|$是$\mathbb{C}^{n\times n}$上的矩阵范数,幂级数

$$f(t)\equiv\sum_{k=0}^{\infty}a_k t^k$$

有收敛半径$r>0$,并定义

$$f_{\mathrm{abs}}(t)\equiv\sum_{k=0}^{\infty}|a_k|\ t^k.$$

设$A,E\in\mathbb{C}^{n\times n}$,$\|A\|+\|E\|<r$,则

$$\begin{aligned}\|f(A+E)-f(A)\|&\le f_{\mathrm{abs}}(\|A\|+\|E\|)-f_{\mathrm{abs}}(\|A\|)\\&\le\|E\|f'_{\mathrm{abs}}(\|A\|+\|E\|),\end{aligned}\qquad(3.27)$$

其中$f(A)$或者依 **9.3.2** 作为准素矩阵函数求值,或者等价地按幂级数

$$f(A)=a_0I+a_1A+a_2A^2+\cdots$$

求值,$f(A+E)$类似.

特别,$f(A)$是

$$\{A\in\mathbb{C}^{n\times n}:\rho(A)<r\}$$

上的连续函数;$f(A)$在

$$\{A\in\mathbb{C}^{n\times n}:\|A\|<r/2\}$$

的每个紧子集K上 Lipschitz 连续,也就是说,对于每个这样的K存

在正数 $L=L(K)$,使得

$$\|f(B)-f(A)\| \le L\|B-A\|, \quad \forall A,B\in K .$$

可取 Lipschitz 常数为

$$L=\max\{f'_{\text{abs}}(\|A\|+\|B-A\|): A,B\in K\}.$$

证 利用(2.16),

$$\begin{aligned}\|f(A+E)-f(A)\| &= \left\|\sum_{k=0}^{\infty}a_k\left[(A+E)^k-A^k\right]\right\| \\ &\le \sum_{k=0}^{\infty}|a_k|\left\|(A+E)^k-A^k\right\| \\ &\le \sum_{k=0}^{\infty}|a_k|\left[(\|A\|+\|E\|)^k-\|A\|^k\right] \\ &= f_{\text{abs}}(\|A\|+\|E\|)-f_{\text{abs}}(\|A\|) \\ &= \|E\|f'_{\text{abs}}(\xi)\le \|E\|f'_{\text{abs}}(\|A\|+\|E\|),\end{aligned}$$

其中 $\|A\|\le\xi\le\|A\|+\|E\|$.

如果 $\rho(A)<r$,依 **2.1.22**,存在 $\mathbb{C}^{n\times n}$ 上的矩阵范数 $\|\bullet\|$ 使得 $\|A\|<r$;又因在 $[0,r]$ 上 f_{abs} 是解析的,故必连续.

于是,当 $E\to 0$ 时

$$f_{\text{abs}}(\|A\|+\|E\|)\to f_{\text{abs}}(\|A\|),$$

并由此推出 $f(A)$ 的连续性.关于 $f(A)$ 的 Lipschitz 连续,可以仿照 **9.2.3** 中的证明. □

这一定理特别简单的应用例子是指数函数 $f(t)=\mathrm{e}^t$.从

$$r=\infty,\ f_{\text{abs}}=f$$

和

$$f(\|A\|+\|E\|)=f(\|A\|)f(\|E\|)$$

得下述推论.

9.3.13 推论 对于任何 $A,E\in\mathbb{C}^{n\times n}$,以及$\mathbb{C}^{n\times n}$上的任何矩阵范数 $\|\bullet\|$,成立

$$\left\|e^{A+E}-e^{A}\right\|\le\left(\exp\|E\|-1\right)\exp\|A\|$$
$$\le\|E\|\exp\|E\|\exp\|A\|. \tag{3.28}$$

特别,函数

$$f(A)\equiv \mathrm{e}^{A}$$

在$\mathbb{C}^{n\times n}$上连续,并在$\mathbb{C}^{n\times n}$的每个紧子集上 Lipschitz 连续. □

9.3.14 定理 设 $A\in\mathbb{C}^{n\times n}$,其最小多项式为

$$m(t)=(t-\lambda_1)^{m_1}\cdots(t-\lambda_s)^{m_s},$$

其中 $\lambda_1,\cdots,\lambda_s$ 相异且 $m_k\ge 1, k=1,\cdots,s$. 设 $t_0\in\mathbb{C}, t_0\ne 0$, f 是 $D\subset\mathbb{C}$上的实或复值函数,且

$$t_0\lambda_1,\cdots,t_0\lambda_s\in D,$$

f 在每点 $t_0\lambda_k$ 和其一个开邻域上分别 m_k 次可微和 m_k-1 次可微,$k=1,\cdots,s$,则

(1) 相伴于函数 f 的准素矩阵函数 $f(tA)$ 对于 $t=t_0$ 的一个开邻域内的所有 t 有定义.

(2) 相伴于函数 f' 的准素矩阵函数 $f'(t_0A)$ 有定义.

(3) $f(tA)$ 在 $t=t_0$ 是 t 的可微函数,而且

$$\left.\frac{\mathrm{d}}{\mathrm{d}t}f(tA)\right|_{t=t_0}=Af'(t_0A)=f'(t_0A)A. \tag{3.29}$$

特别,对于 $f(t)\equiv\mathrm{e}^{t}$,有

$$\frac{\mathrm{d}}{\mathrm{d}t}\mathrm{e}^{tA}=A\mathrm{e}^{tA}=\mathrm{e}^{tA}A,\quad \forall A\in\mathbb{C}^{n\times n}. \tag{3.30}$$

证 可微性和定义域的假设足以在 t_0 一个邻域内对 t 定义准素矩阵函数 $f(tA)$ 及定义 $f'(t_0A)$.

设 $A=SJS^{-1}$, J 是 A 的 Jordan 标准形,形如(3.11),则

$$f(tA)=f\left(S(tJ)S^{-1}\right)=Sf(tJ)S^{-1}$$
$$=S\begin{bmatrix} f\left(tJ_{n_1}\left(\lambda_{s_1}\right)\right) & & 0 \\ & \ddots & \\ 0 & & f\left(tJ_{n_r}\left(\lambda_{s_r}\right)\right) \end{bmatrix}S^{-1}.$$

故只须证明单个 Jordan 块成立(3.29).

设 $J_k(\lambda)$ 是 A 的 Jordan 标准形中一个 $k\times k$ 的 Jordan 块.依定理假设,f 在 $t=t_0\lambda$ 至少 k 次可微,在 $t_0\lambda$ 的一个邻域内至少 $k-1$ 次可微.对于 $t\neq 0$,令

$$S_t\equiv \operatorname{diag}\left(1,\frac{1}{t},\cdots,\frac{1}{t^{k-1}}\right)\in\mathbb{C}^{k\times k}.$$

容易看出,$tJ_k(\lambda)=S_tJ_k(t\lambda)S_t^{-1}$,从而

$$f\left(tJ_k(\lambda)\right)=S_tf\left(J_k(t\lambda)\right)S_t^{-1}.$$

根据(3.13),这是一个上三角 Toeplitz 矩阵,其 $(1,j)$-元素是

$$\frac{t^{j-1}f^{(j-1)}(t\lambda)}{(j-1)!},\quad j=1,\cdots,k.$$

$f\left(tJ_k(\lambda)\right)$ 采用如此表示,对于 $t=0$ 也是有效的,因为此时它给出的是 $f(0)=f(0)I$ (注意,这里等式左端的 0 是零矩阵,右端的 0 是数零).用 f' 替换 f,得知 $f'\left(t_0J_k(\lambda)\right)$ 是上三角 Toeplitz 矩阵,其 $(1,j)$-元素是

$$\frac{t_0^{j-1}f^{(j-1)}(t_0\lambda)}{(j-1)!},\quad j=1,\cdots,k.$$

现在直接求导,

$$\left.\frac{\mathrm{d}}{\mathrm{d}t}f\left(tJ_k(\lambda)\right)\right|_{t=t_0}$$

是上三角 Toeplitz 矩阵,其 $(1,1)$-元素是

$$\lambda f'(t_0\lambda),$$

$(1,j)$-元素是

$$\frac{t_0^{j-2}f^{(j-1)}(t_0\lambda)}{(j-2)!}+\lambda\frac{t_0^{j-1}f^{(j)}(t_0\lambda)}{(j-1)!},\quad j=2,\cdots,k.$$

这些正好也是上三角 Toeplitz 矩阵 $f'(t_0J_k(\lambda))J_k(\lambda)$ 的第1行的元素,因此

$$\frac{\mathrm{d}}{\mathrm{d}t}f(tJ_k(\lambda))\bigg|_{t=t_0}=f'(t_0J_k(\lambda))J_k(\lambda).$$

可交换的结论 $f'(t_0A)A=Af'(t_0A)$ 直接从准素矩阵函数 $f'(t_0A)$ 是 t_0A 的多项式推出. □

重要公式(3.30)表明:对于任何 $A\in\mathbb{C}^{n\times n}$ 及 $x(0)\in\mathbb{C}^n$, 一阶向量初值问题

$$\begin{cases}\dfrac{\mathrm{d}}{\mathrm{d}t}x(t)=Ax(t),\\ x(t)\big|_{t=0}=x(0)\end{cases}\tag{3.31}$$

的解是 $x(t)=\mathrm{e}^{tA}x(0)$.微分方程一般理论保证此问题有唯一解.由这一事实可以引出计算关于任何 $A\in\mathbb{C}^{n\times n}$ 的 e^{tA} 的避免幂级数及 Jordan 标准形的简单方法.

9.3.15 定理 设 $D\subset\mathbb{C}$,集合

$$\mathscr{N}_n(D)\equiv\{A\in\mathbb{C}^{n\times n}:A\text{ 正规且 }\lambda(A)\subset D\},$$

f 是 D 上连续的实或复值函数,则准素矩阵函数

$$f(A)=U\begin{bmatrix}f(\lambda_1) & & 0\\ & \ddots & \\ 0 & & f(\lambda_n)\end{bmatrix}U^*\tag{3.32}$$

在 $\mathscr{N}_n(D)$ 连续,其中

$$A=U\Lambda U^*,\quad \Lambda\equiv\operatorname{diag}(\lambda_1,\cdots,\lambda_n),\tag{3.33}$$

$U \in \mathbb{C}^{n\times n}$ 是酉矩阵.

证 正规矩阵是酉可对角化的,因此对于给定的 $A \in \mathscr{N}_n(D)$,可设其表示为(3.33).设 $p(t)$ 是多项式,使得

$$p(\lambda_i) = f(\lambda_i), \quad i = 1,\cdots,n,$$

依 **9.3.2**, $f(A) = p(A)$.

现在,任给 $\varepsilon > 0$ 选取 $\delta > 0$ 充分小,使得

(1) $\|p(A) - p(B)\|_2 \le \varepsilon/3$,当 $B \in \mathbb{C}^{n\times n}$ 且 $\|A - B\|_2 < \delta$(利用 **9.2.3**);

(2) $|p(t) - p(\lambda_i)| < \varepsilon/3$,当 $|t - \lambda_i| \le \delta$, $i = 1,\cdots,n$;

(3) $|f(t) - f(\lambda_i)| < \varepsilon/3$,当 $t \in D$, $|t - \lambda_i| \le \delta$, $i = 1,\cdots,n$.

设 $B \in \mathscr{N}_n(D)$ 满足 $\|A - B\|_2 \le \delta$.根据 Hoffman-Wielandt 定理(见[HJ1985]中的定理 6.3.5),存在酉矩阵 $V \in \mathbb{C}^{n\times n}$,使得

$$B = VMV^*, \quad M = \operatorname{diag}(\mu_1,\cdots,\mu_n);$$

$$|\mu_i - \lambda_i| \le \delta, \ i = 1,\cdots,n.$$

利用谱范数 $\|\bullet\|$ 的酉不变性,有

$$\begin{aligned}
\|f(A) - f(B)\|_2 &= \|p(A) - p(B) + p(B) - f(B)\|_2 \\
&\le \|p(A) - p(B)\|_2 + \|p(B) - f(B)\|_2 \\
&= \|p(A) - p(B)\|_2 + \|V(p(M) - f(M))V^*\|_2 \\
&\le \|p(A) - p(B)\|_2 + \|p(M) - f(\Lambda)\|_2 + \|p(\Lambda) - f(M)\|_2 \\
&= \|p(A) - p(B)\|_2 + \max_{1\le i\le n}|p(\mu_i) - p(\lambda_i)| \\
&\quad + \max_{1\le i\le n}|f(\lambda_i) - f(\mu_i)| \\
&\le \frac{\varepsilon}{3} + \frac{\varepsilon}{3} + \frac{\varepsilon}{3} = \varepsilon.
\end{aligned}$$

□

9.4 Hadamard 矩阵函数

9.4.1 定义 设有n^2个定义在$\Omega \subset \mathbb{C}$的函数

$$f_{ij}, \quad i,j=1,\cdots,n, \tag{4.1}$$

映射$f:\Omega^{n\times n} \to \mathbb{C}^{n\times n}$,

$$f(A)=\left[f_{ij}(a_{ij})\right], \quad \forall A=[a_{ij}]\in \Omega^{n\times n}, \tag{4.2}$$

则称f为**Hadamard(矩阵)函数**.

特别,若

$$f_{ij}(t)=t^{\alpha}, \qquad i,j=1,\cdots,n,$$

其中$\alpha \in \mathbb{C}$是常数,记

$$A^{(\alpha)} \equiv f(A)=\left[a_{ij}^{\alpha}\right], \quad \forall A=[a_{ij}]\in \mathbb{C}^{n\times n}, \tag{4.3}$$

则称$A^{(\alpha)}$是**Hadamard 幂**.

如果每个函数f_{ij}是多项式,

$$f_{ij}(t)=c_{k_{ij}}(i,j)t^{k_{ij}}+c_{k_{ij}-1}(i,j)t^{k_{ij}-1}+\cdots+c_0(i,j),$$
$$i,j=1,\cdots,n, \tag{4.4}$$

那么$f(A)=\left[f_{ij}(a_{ij})\right]$可以写成

$$f(A)=\left(C_m \circ A^{(m)}\right)+\left(C_{m-1} \circ A^{(m-1)}\right)+\cdots+C_0, \tag{4.5}$$

其中

$$m=\max\left\{k_{ij}: i,j=1,\cdots,n\right\}$$

及

$$C_k=\left[c_k(i,j)\right], \quad k=0,1,\cdots,m,$$

而$A^{(k)}=\left[a_{ij}^{k}\right]$是 Hadamard 幂,$C_k \circ A^{(k)}$是 Hadamard 积.

如果每个函数 f_{ij} 是解析函数,f_{ij} 的幂级数展开具有收敛半径 $r_{ij}>0, i,j=1,\cdots,n$,那么 $f(A)=[f_{ij}(a_{ij})]$ 可表示成幂级数

$$f(A)=\sum_{k=0}^{\infty}c_k\circ A^{(k)},$$

$$\forall A=[a_{ij}]\in\mathbb{C}^{n\times n},\ |a_{ij}|<r_{ij},\ i,j=1,\cdots,n, \quad (4.6)$$

其中所有 $A^{(k)}$ 是 Hadamard 幂,所有 $C_k\circ A^{(k)}$ 是 Hadamard 积.

记号上和概念上,(4.5)和(4.6)的最简单情形是所有函数 f_{ij} 均相同的情形,记之为 f.此时,可直接将 f 用作矩阵函数的记号:

$$f(A)=[f(a_{ij})].$$

如果 $f(t)=\sum_{k=0}^{\infty}a_k t^k$,那么(4.6)成为

$$f(A)=\sum_{k=0}^{\infty}(a_k J)\circ A^{(k)}=\sum_{k=0}^{\infty}a_k A^{(k)}, \quad (4.7)$$

其中 $J\in\mathbb{C}^{n\times n}$ 是所有元素为1的矩阵,这是关于 Hadamard 乘法的"单位矩阵".

例 设 $f(t)=\mathrm{e}^t$, 则

$$f(A)=[\mathrm{e}^{a_{ij}}]=\sum_{k=0}^{\infty}\frac{1}{k!}A^{(k)}=\sum_{k=0}^{\infty}\frac{1}{k!}[a_{ij}^k],$$

$$\forall A=[a_{ij}]\in\mathbb{C}^{n\times n}. \quad (4.8)$$

矩阵的 Hadamard 函数自然地出现在积分算子、积分方程和整二次型的离散化的研究之中.因为通常在物理问题中遇到的是正定积分算子,Hadamard 矩阵函数的许多已有结果是有关正定矩阵的 Hadamard 函数的.

9.4.2 定理 设 $A\in\mathbb{C}^{n\times n}$ 是半正定矩阵.若(4.6)中所有系数矩阵 C_k 是半正定的,则由(4.6)定义的 Hadamard 函数 f 是半正定的.

特别,若(4.7)中所有系数 a_k 是非负的,则由(4.7)定义的 Hadamard 函数 f 是半正定的.

证 依假设,Schur 积定理 **7.2.9** 保证每个整数次 Hadamard 幂 $A^{(k)}$ 是半正定的,而且每个 Hadamard 积 $C_k\circ A^{(k)}$ 是半正定的.然后,从半正定矩阵之和必为半正定矩阵即得结论. □

由这个定理及(4.8)推出,当 $A=[a_{ij}]$ 为半正定时, Hadamard 指数

$$e^{\circ A}\equiv\left[e^{a_{ij}}\right] \tag{4.9}$$

是半正定的.

实际上,针对指数函数这一特殊情形,还可以得到稍强一些的结果.

9.4.3 定义 设 $A\in\mathbb{C}^{n\times n}$ 是自伴矩阵,且满足性质

$$x^*Ax\geq 0,\quad \forall x=\left(x_1,\cdots,x_n\right)^{\mathrm{T}}\in\mathbb{C}^n, x_1+\cdots+x_n=0, \tag{4.10}$$

则称 A 为**条件半正定矩阵**(conditionally positive semidefinite matrix),有时也称为**几乎半正定矩阵**(almost positive semidefinite matrix).

9.4.4 定理 设 $A=[a_{ij}]\in\mathbb{C}^{n\times n}$ 是条件半正定矩阵,则 Hadamard 指数 $e^{\circ A}\equiv\left[e^{a_{ij}}\right]$ 是半正定矩阵.

证 令 $e=\left(1,\cdots,1\right)^{\mathrm{T}}\in\mathbb{C}^n$,对于任一 $x\in\mathbb{C}^n$,向量

$$\hat{x}\equiv x-\frac{1}{n}ee^*x$$

满足分量之和 $e^*\hat{x}=0$;因此

$$0 \le \hat{x}^* A \hat{x} = \left(x - \frac{1}{n} ee^* x\right)^* A\left(x - \frac{1}{n} ee^* x\right)$$

$$= x^* \left\{A - \frac{1}{n} ee^* A - \frac{1}{n} Aee^* + \left(\frac{1}{n}\right)^2 \left(e^* Ae\right) ee^* \right\} x.$$

这说明矩阵

$$B \equiv A - \frac{1}{n} ee^* A - \frac{1}{n} Aee^* + \left(\frac{1}{n}\right)^2 \left(e^* Ae\right) ee^* = [b_{ij}]$$

是半正定的.令向量

$$y \equiv \frac{1}{n} Ae - \frac{1}{2}\left(\frac{1}{n}\right)^2 \left(e^* Ae\right) e = \left(y_1, \cdots, y_n\right)^{\mathrm{T}},$$

则有

$$A = B + ye^* + ey^*$$

或

$$a_{ij} = b_{ij} + y_i + \overline{y_j}, \quad i, j = 1, \cdots, n.$$

于是

$$\mathrm{e}^{\circ A} = \left[\mathrm{e}^{b_{ij} + y_i + \overline{y_j}}\right] = \left[\mathrm{e}^{b_{ij}} \mathrm{e}^{y_i} \mathrm{e}^{\overline{y_j}}\right].$$

由此,并因 B 半正定,依 **9.4.2**,

$$z^* \mathrm{e}^{\circ A} z = \sum_{i,j=1}^{n} \mathrm{e}^{b_{ij}} \left(z_i \mathrm{e}^{y_i}\right) \overline{\left(z_j \mathrm{e}^{y_j}\right)} \ge 0,$$

$$\forall z = \left(z_1, \cdots, z_n\right)^{\mathrm{T}} \in \mathbb{C}^n.$$

□

9.4.5 定理 设 f 是区间 $(0, \infty)$ 上 $n-1$ 次连续可微的实值函数,而且对于每个具有正元素的半正定矩阵 $A = [a_{ij}] \in \mathbb{C}^{n \times n}$,矩阵 $f(A) \equiv \left[f(a_{ij})\right]$ 是半正定的,则

$$f^{(k)}(t) \ge 0, \ \forall t \in (0, \infty); \quad k = 0, 1, \cdots, n-1. \quad (4.11)$$

证 应用数学归纳法.

当 $n=1$ 时,定理的假设明显保证 $f(t)\geq 0,\forall t>0$.

考虑 $n\geq 2$.定理的假设保证当 $B\in\mathbb{C}^{m\times m}$ 半正定且具有正元素时,$f(B)$ 是半正定的,$1\leq m\leq n$.于是,可以假定

$$f^{(k)}(t)\geq 0,\ \ \forall t\in(0,\infty);\quad k=0,1,\cdots,n-2, \tag{4.12}$$

要证明 $f^{(n-1)}(t)\geq 0,\forall t>0$.

设 $a>0,\alpha=(\alpha_1,\cdots,\alpha_n)^{\mathrm{T}}\in\mathbb{R}^n$,使得对于充分小的 $t>0$,矩阵 $A(t)\equiv[a+t\alpha_i\alpha_j]$ 是半正定且具有正元素的.此时,

$$f(A(t))=[f(a+t\alpha_i\alpha_j)]$$

是半正定的,从而

$$\Delta(t)\equiv\det f(A(t))\geq 0,$$

因为 $f(A(0))$ 的所有元素为 a,$\Delta(0)=0$.而且

$$\left.\frac{\mathrm{d}^k}{\mathrm{d}t^k}\Delta(t)\right|_{t=0}\equiv\Delta^{(k)}(0)=0,\quad k=0,1,\cdots,\frac{n(n-1)}{2}-1,$$

原因是从 k 次微分产生的 n^k 个行列式项的每个行列式中至少有两列成比例.令 $\eta=n(n-1)/2$,通过直接计算可得

$$0\leq\lim_{t\to 0}\frac{\Delta(t)}{t^\eta}=\Delta^{(\eta)}(0)$$

$$=\binom{\eta}{1}\binom{\eta-1}{2}\binom{\eta-3}{3}\binom{\eta-6}{4}\cdots\binom{2n-3}{n-2}V^2(\alpha)\prod_{k=0}^{n-1}f^{(k)}(a), \tag{4.13}$$

其中 $\binom{m}{l}$ 是二项式系数,而

$$V(\alpha)\equiv\det\begin{bmatrix}1 & \alpha_1 & \alpha_1^2 & \cdots & \alpha_1^{n-1}\\ 1 & \alpha_2 & \alpha_2^2 & \cdots & \alpha_2^{n-1}\\ \cdots & \cdots & \cdots & \cdots & \cdots\\ 1 & \alpha_n & \alpha_n^2 & \cdots & \alpha_n^{n-1}\end{bmatrix}=\prod_{1\leq j<i\leq n}(\alpha_i-\alpha_j)$$

是 Vandermonde 行列式.注意,当$\alpha_1,\cdots,\alpha_n$互不相等时,

$$V^2(\alpha)>0.$$

于是,从(4.13)得

$$f(a)f'(a)\cdots f^{(n-1)}(a)\geq 0. \tag{4.14}$$

然而,由此及归纳法假设(4.12)下述不能推出$f^{(n-1)}(a)\geq 0$.

现在,考虑

$$g(t)\equiv f(t)+\tau t^n,$$

显然,对于所有$\tau\geq 0$,g保有f所具有的性质.因此,对(4.14)用g代替f,有

$$p(\tau)\equiv\left(f(a)+\tau a^n\right)\left(f'(a)+\tau n a^{n-1}\right)\cdots\left(f^{(n-1)}(a)+\tau n! a\right)\geq 0,\quad \forall\tau\geq 0.$$

由于$p(\tau)$是至多有n个实零点的多项式,所以必有$\varepsilon>0$,使得当$0<\tau<\varepsilon$时$p(\tau)>0$.因此,从归纳法假设(4.12)得

$$f^{(n-1)}(a)+\tau n! a>0,\quad 0<\tau<\varepsilon,$$

从而有$f^{(n-1)}(a)\geq 0$,$\forall a>0$. □

9.4.6 推论 设$0<\alpha<n-2$,α非整数,则存在具有正元素的半正定矩阵$A=[a_{ij}]\in\mathbb{C}^{n\times n}$,使得 Hadamard 幂$A^{(\alpha)}\equiv\left[a_{ij}^{\alpha}\right]$是非半正定的.

证 考虑$f(t)=t^{\alpha}$.当$\alpha<n-2$不是正整数时,导数

$$f'(t)=\alpha t^{\alpha-1},$$

$$\cdots$$

$$f^{(n-1)}(t)=\alpha(\alpha-1)\cdots(\alpha-n+2)t^{\alpha-n+1},$$

对于$t>0$不可能全是非负的.f不符合 **9.4.5** 的必要条件,从而必存在具有正元素的半正定矩阵$A=[a_{ij}]$,使得

$$f(A)\equiv\left[f(a_{ij})\right]=\left[a_{ij}^{(\alpha)}\right]\equiv A^{(\alpha)}$$

是非半正定的. □

9.4.7 定理 设 $A=[a_{ij}]\in\mathbb{C}^{n\times n}$ 是具有正元素的半正定矩阵,如果 $\alpha\geq n-2$,那么 Hadamard 幂 $A^{(\alpha)}\equiv\left[a_{ij}^{\alpha}\right]$ 是半正定的.而且,一般地说,$n-2$ 是最好可能的下界.

证 如果 $n=2$,且半正定矩阵

$$A=\begin{bmatrix}a_{11} & a_{12}\\ a_{12} & a_{22}\end{bmatrix}$$

是非奇异的,那么必有

$$a_{11}>0,\qquad \det A=a_{11}a_{22}-a_{12}^{2}>0 \text{ 或 } a_{11}a_{22}>a_{12}^{2}.$$

因此,对任何 $\alpha>0$,仍有

$$a_{11}^{\alpha}>0,\qquad a_{11}^{\alpha}a_{22}^{\alpha}>\left(a_{12}^{2}\right)^{\alpha} \text{ 或 } a_{11}^{\alpha}a_{22}^{\alpha}-\left(a_{12}^{2}\right)^{\alpha}>0.$$

依 **3.10.6**,对所有 $\alpha>n-2=2-2=0$,$A^{(\alpha)}$ 是正定的.对于 A 是奇异的情形,可以利用非奇异情形取极限推出 $A^{(\alpha)}$ 是半正定的.

现在设 $n\geq 3$,而且假设定理对所有不超过 $n-1$ 的满足假定的矩阵成立.将 A 分块写成

$$A=\begin{bmatrix}B & \omega\\ \omega^{\mathrm{T}} & a_{nn}\end{bmatrix},$$

其中

$$B=\begin{bmatrix}a_{11} & a_{12} & \cdots & a_{1,n-1}\\ a_{12} & \ddots & \ddots & \vdots\\ \vdots & \ddots & \ddots & a_{n-2,n-1}\\ a_{1,n-1} & \cdots & a_{n-2,n-1} & a_{n-1,n-1}\end{bmatrix},\quad \omega=\begin{pmatrix}a_{1n}\\ a_{2n}\\ \vdots\\ a_{n-1,n}\end{pmatrix}.$$

设 $\eta^{\mathrm{T}}=\left(\xi^{\mathrm{T}},z\right)$,$\xi\in\mathbb{C}^{n-1}$,$z\in\mathbb{C}$,则因 A 是半正定的,

$$\eta^{*}A\eta=\xi^{*}B\xi+2\operatorname{Re}\left(z\xi^{*}\omega\right)+a_{nn}|z|^{2}\geq 0,$$

$$\forall\xi\in\mathbb{C}^{n-1},\ z\in\mathbb{C}.\qquad(4.15)$$

如果 $a_{nn}=0$,那么只有 $\omega=0$ 时(4.15)才能成立,而此时 A 的最后

的行和列为零,根据归纳法假设,对所有$\alpha \ge (n-1)-2=n-3$,矩阵$A^{(\alpha)}$是半正定的.

可以假定$a_{nn}>0$.在(4.15)中取

$$z=-\omega^{\mathrm T}\xi/a_{nn},$$

得

$$\eta^* A\eta=\xi^*\left(B-\omega\omega^{\mathrm T}/a_{nn}\right)\xi\ge 0,\quad \forall \xi\in\mathbb{C}^{n-1},$$

推出$B-\omega\omega^{\mathrm T}/a_{nn}$是半正定的.注意,若取

$$\zeta\equiv\left(\omega^{\mathrm T},a_{nn}\right)^{\mathrm T}/\sqrt{a_{nn}}=\left(a_{1n},a_{2n},\cdots,a_{nn}\right)^{\mathrm T}/\sqrt{a_{nn}},$$

则

$$A-\zeta\zeta^{\mathrm T}=\begin{bmatrix} B-\omega\omega^{\mathrm T}/a_{nn} & 0\\ 0 & 0\end{bmatrix} \tag{4.16}$$

是半正定的.

引进积分恒等式

$$a^{\alpha}=c^{\alpha}+\alpha\int_0^1(a-c)\left(ta+(1-t)c\right)^{\alpha-1}\mathrm{d}t,$$
$$\forall \alpha\ge 1, a\ge 0, c\ge 0,$$

按逐个元素应用此恒等式得出矩阵恒等式

$$\begin{aligned}\left[a_{ij}^{\alpha}\right]&\equiv A^{(\alpha)}\\ &=\left(\zeta\zeta^{\mathrm T}\right)^{(\alpha)}+\alpha\int_0^1\left(A-\zeta\zeta^{\mathrm T}\right)\circ\left(tA+(1-t)\zeta\zeta^{\mathrm T}\right)^{(\alpha-1)}\mathrm{d}t,\end{aligned}$$
$$\forall \alpha\ge 1, \tag{4.17}$$

其中被积函数是$A-\zeta\zeta^{\mathrm T}$和$\left(tA+(1-t)\zeta\zeta^{\mathrm T}\right)^{(\alpha-1)}$的 Hadamard 积.因$A-\zeta\zeta^{\mathrm T}$的最后的行和列为零,而$tA+(1-t)\zeta\zeta^{\mathrm T}$的左上$n-1$维主子矩阵的$\alpha-1$次 Hadamard 幂,根据归纳法假设当

$$\alpha-1\ge(n-1)-2=n-3 \quad 即 \quad \alpha\ge n-2$$

时是半正定的,故依 Schur 积定理 **7.2.9**,(4.17)的被积函数是半正定的.这样,积分项是以半正定矩阵非负线性组合表示的 Riemann 和的

极限,从而是半正定的;而(4.17)的第一项

$$\left(\zeta\zeta^{\mathrm{T}}\right)^{(\alpha)}=\left[\zeta_i^{\alpha}\zeta_j^{\alpha}\right];\quad \zeta_i=a_{in}\big/\sqrt{a_{nn}}\,,i=1,\cdots,n$$

显然是半正定的.因此,当$\alpha\ge n-2$时$A^{(\alpha)}$是半正定的.

推论 **9.4.6** 表明$n-2$是最好可能的下界. □

然而, $\alpha\ge n-2$并不是$A^{(\alpha)}$半正定的必要条件.一个典型的例子是:若$A\in\mathbb{C}^{n\times n}$是条件半正定矩阵,且$\alpha>0$,则$\alpha A$是条件半正定的,从而依 **9.4.4**,对所有$\alpha>0$, Hadamard 指数

$$e^{\circ\alpha A}=\left[e^{\alpha a_{ij}}\right]=\left[e^{a_{ij}}\right]^{(\alpha)}$$

是半正定的.

9.4.8 定义 设$A\in\mathbb{C}^{n\times n}$是半正定矩阵,而且对所有$\alpha>0$, $A^{(\alpha)}$是半正定的,则称A为**无穷可分矩阵**(infinitely divisible matrix).

这一术语来自概率论中的**无穷可分特征函数**理论.在那里是指连续函数$\phi:\mathbb{R}\to\mathbb{C}$,满足$\phi(0)=1$,且对所有$\alpha>0$,核

$$K^{\alpha}(s,t)=\phi^{\alpha}(s-t)$$

是半正定的.

9.4.9 定理 设$A\in\mathbb{R}^{n\times n}$是对称正矩阵,则$A$无穷可分的充分必要条件是 Hadamard 对数$\log^{\circ}A\equiv\left[\log a_{ij}\right]$为条件半正定.

证 用$J\in\mathbb{C}^{n\times n}$表示所有元素为1的矩阵.若$A^{(\alpha)}$对所有$\alpha>0$是半正定的,而且$x=(x_1,\cdots,x_n)^{\mathrm{T}}\in\mathbb{C}^n$, $x_1+\cdots+x_n=0$,则有

$$x^*\frac{1}{\alpha}A^{(\alpha)}x=x^*\left(\frac{1}{\alpha}A^{(\alpha)}-\frac{1}{\alpha}J\right)x=x^*\left(\frac{1}{\alpha}\left(A^{(\alpha)}-J\right)\right)x\ge 0.$$

由此,令$\alpha\to 0$,并利用

$$\lim_{\alpha\to 0}\frac{a_{ij}^{\alpha}-1}{\alpha}=\log a_{ij},\quad i,j=1,\cdots,n,$$

即得 $\log^{\circ} A$ 是条件半正定的.必要性得证.

关于充分性,利用 **9.4.4**,得知

$$A^{(\alpha)}=\left[e^{\alpha\log a_{ij}}\right],$$

对 $\alpha>0$ 是半正定的. □

9.5 函数矩阵

9.5.1 引言 现在转向函数的矩阵,包括**向量值函数**

$$x:D\subset\mathbb{C}\to\mathbb{C}^n,\quad x(t)=\left(x_1(t),\cdots,x_n(t)\right)^{\mathrm{T}}$$

和**矩阵值函数**

$$A:D\subset\mathbb{C}\to\mathbb{C}^{n\times n},\quad A(t)=\left[a_{ij}(t)\right].$$

本节主要引进函数矩阵的微分学.

存在两种途径:一种是按逐个分量或元素的通常极限、连续和导数等建立函数矩阵的相应概念并引出种种性质;另一种是利用赋范线性空间中的向量范数或矩阵范数,随之定义向量序列或矩阵序列的收敛性,重新建立理论.

这里选择第二种途径,原因是它既简单,又能提供关于课题思索概念形成的模式;但在必要时仍采用分量或元素的说法.

注意这一节讨论的矩阵值函数和前一节讨论的 Hadamard 函数的区别.

下面首先引入矩阵序列收敛的定义,它涵盖了 **1.7.10** 已经给出的向量序列收敛的定义.

9.5.2 定义 设 $\|\cdot\|$ 是 $\mathbb{C}^{n\times n}$ 上的范数.矩阵序列 $\{A_k\}\subset\mathbb{C}^{n\times n}$ 称为**收敛于极限** $A\in\mathbb{C}^{n\times n}$,记作 $\lim\limits_{k\to\infty}A_k=A$,如果

$$\lim_{k\to\infty}\|A_k - A\| = 0.$$

在 **1.7.10** 和这里定义的向量序列收敛和矩阵序列收敛,没有采用要求向量序列 $\{x^{(k)}\}$ 按分量趋向极限 x 的相应分量和矩阵序列 $\{A_k\}$ 按元素趋向极限 A 的相应分量元素的说法.但这种说法可以让概念和思考得以简化,并且在论证上也是有用的.

9.5.3 定义 设 $x: D\subset\mathbb{C}\to\mathbb{C}^n$ 是向量值函数,$\|\bullet\|$ 是 $\mathbb{C}^n$ 上的范数. x 称为在 $t_0\in D$ **连续**,如果

$$\lim_{t\to t_0}\|x(t)-x(t_0)\| = 0. \tag{5.1}$$

x 称为在 $t_0\in D$ **可微**,具有**导数** $\dot{x}(t_0)\in\mathbb{C}^n$,如果

$$\lim_{h\to 0}\left\|\frac{x(t_0+h)-x(t_0)}{h}-\dot{x}(t_0)\right\| = 0. \tag{5.2}$$

类似地,设 $A: D\subset\mathbb{C}\to\mathbb{C}^{n\times n}$ 是矩阵值函数,$\|\bullet\|$ 是 $\mathbb{C}^{n\times n}$ 上的范数. A 称为在 $t_0\in D$ **连续**,如果

$$\lim_{t\to t_0}\|A(t)-A(t_0)\| = 0.$$

A 称为在 $t_0\in D$ **可微**,具有**导数** $\dot{A}(t_0)\in\mathbb{C}^n$,如果

$$\lim_{h\to 0}\left\|\frac{A(t_0+h)-A(t_0)}{h}-\dot{A}(t_0)\right\| = 0.$$

这里用一点来缩记导数

$$\dot{x}(t) = \frac{\mathrm{d}}{\mathrm{d}t}x(t),\quad \dot{A}(t) = \frac{\mathrm{d}}{\mathrm{d}t}A(t).$$

下面的定理表明微分学基本引理对向量值函数和矩阵值函数成立.

9.5.4 定理 设 $x:(a,b)\subset\mathbb{R}\to\mathbb{R}^n$ 是向量值函数,而且

$$\dot{x}(t)=0,\quad \forall t\in(a,b),$$

则 $x(t)$ 在 (a,b) 上为常向量.

证 根据假设,

$$\frac{\mathrm{d}}{\mathrm{d}t}(x(t),y)=(\dot{x}(t),y)=0,\quad \forall t\in(a,b),y\in\mathbb{R}^n,$$

这里 $(\bullet,\bullet)$ 是 $\mathbb{R}^n$ 上的 Euclid 内积.于是,根据一元微分学基本引理,在区间 (a,b) 上,对于每个 $y\in\mathbb{R}^n$,$(x(t),y)$ 是常数.特别,对任一固定的 $t_0\in(a,b)$,

$$(x(t),y)=(x(t_0),y),\quad \forall t\in(a,b),y\in\mathbb{R}^n,$$

这蕴涵

$$x(t)=x(t_0),\ \forall t\in(a,b).$$

□

对于矩阵值函数有同样的结果.

下述几个定理给出向量值函数和矩阵值函数的基本微分法则.

9.5.5 定理 线性微分法则:

(1) 两个可微向量值函数 x 与 y 之和是可微的,而且

$$\frac{\mathrm{d}}{\mathrm{d}t}(x(t)+y(t))=\frac{\mathrm{d}}{\mathrm{d}t}x(t)+\frac{\mathrm{d}}{\mathrm{d}t}y(t).$$

两个可微矩阵值函数 A 与 B 之和是可微的,而且

$$\frac{\mathrm{d}}{\mathrm{d}t}(A(t)+B(t))=\frac{\mathrm{d}}{\mathrm{d}t}A(t)+\frac{\mathrm{d}}{\mathrm{d}t}B(t).$$

(2) 可微向量值函数 x 与任一常数 k 的数积是可微的,而且

$$\frac{\mathrm{d}}{\mathrm{d}t}(kx(t))=k\frac{\mathrm{d}}{\mathrm{d}t}x(t).$$

可微矩阵值函数 A 与任一常数 k 的数积是可微的,而且

$$\frac{\mathrm{d}}{\mathrm{d}t}\left(kA(t)\right)=k\frac{\mathrm{d}}{\mathrm{d}t}A(t).$$

(3) 设 l 是定义在 $\mathbb{R}^n$ 上的线性函数, $x(\cdot)$ 是可微向量值函数,则 $l(x(\cdot))$ 是可微的,而且

$$\frac{\mathrm{d}}{\mathrm{d}t}l(x(t))=l\left(\frac{\mathrm{d}}{\mathrm{d}t}x(t)\right).$$

(4) 设 A 是可微矩阵值函数,则

$$\frac{\mathrm{d}}{\mathrm{d}t}\mathrm{tr}\left(A(t)\right)=\mathrm{tr}\left(\frac{\mathrm{d}}{\mathrm{d}t}A(t)\right).$$

证 (1)和(2)可仿照一元微分学中的相应证明.

(3)的证明可以利用线性函数 l 能表示成

$$l(x)=(x,y),$$

其中 y 是依赖 l 的固定向量, $(\cdot,\cdot)$ 是向量的 Euclid 内积.

(4)是(3)的一个具体应用,因为矩阵的迹是线性函数. □

9.5.6 定理 积的微分法则:

(1) 可微纯量函数 k 与可微向量值函数 x 之积是可微的,而且

$$\frac{\mathrm{d}}{\mathrm{d}t}\left[k(t)x(t)\right]=\left[\frac{\mathrm{d}}{\mathrm{d}t}k(t)\right]x(t)+k(t)\left[\frac{\mathrm{d}}{\mathrm{d}t}x(t)\right].$$

(2) 可微矩阵值函数 A 与可微向量值函数 x 之积是可微的,而且

$$\frac{\mathrm{d}}{\mathrm{d}t}\left[A(t)x(t)\right]=\left[\frac{\mathrm{d}}{\mathrm{d}t}A(t)\right]x(t)+A(t)\left[\frac{\mathrm{d}}{\mathrm{d}t}x(t)\right].$$

(3) 两个可微矩阵值函数 A 与 B 之积是可微的,而且

$$\frac{\mathrm{d}}{\mathrm{d}t}\left[A(t)B(t)\right]=\left[\frac{\mathrm{d}}{\mathrm{d}t}A(t)\right]B(t)+A(t)\left[\frac{\mathrm{d}}{\mathrm{d}t}B(t)\right].$$

(4) 可微纯量函数 k 与可微矩阵值函数 A 之积是可微的,而且

$$\frac{\mathrm{d}}{\mathrm{d}t}\left[k(t)A(t)\right]=\left[\frac{\mathrm{d}}{\mathrm{d}t}k(t)\right]A(t)+k(t)\left[\frac{\mathrm{d}}{\mathrm{d}t}A(t)\right].$$

(5) 两个可微向量值函数 x 与 y 的内积是可微的,而且

$$\frac{\mathrm{d}}{\mathrm{d}t}\left(x(t),y(t)\right)=\left(\frac{\mathrm{d}}{\mathrm{d}t}x(t),y(t)\right)+\left(x(t),\frac{\mathrm{d}}{\mathrm{d}t}y(t)\right).$$

定理中各条结论的证明与通常纯量函数情形是一样的.它们有时统称为 **Leibniz 法则**.

9.5.7 定理 设 A 是可微和可逆的矩阵值函数, 则 A^{-1} 也是可微的, 而且

$$\frac{\mathrm{d}}{\mathrm{d}t}A^{-1}=-A^{-1}\left(\frac{\mathrm{d}}{\mathrm{d}t}A\right)A^{-1}. \tag{5.3}$$

证 由于

$$A^{-1}(t+h)-A^{-1}(t)=A^{-1}(t+h)\left[A(t)-A(t+h)\right]A^{-1}(t),$$

等式两边除以 h 且让 $h\to 0$ 即得(5.3). □

这一定理表明,阵值函数之逆的微分法则与通常逆函数的微分法则存在微妙的差别.

链微分法则是说:若 f 与 u 是可微纯量函数,则当 f 的定义域包含 u 的值域时,复合函数 $f(u)$ 也是可微的,而且

$$\frac{\mathrm{d}}{\mathrm{d}t}f\left(u(t)\right)=\left[\frac{\mathrm{d}}{\mathrm{d}u}f(u)\right]\left[\frac{\mathrm{d}}{\mathrm{d}t}u(t)\right], \tag{5.4}$$

然而这一法则对矩阵值函数不成立.

例如, $f(u)=u^2$,依 **9.5.6**,

$$\frac{\mathrm{d}}{\mathrm{d}t}\left[A(t)\right]^2=A(t)\left[\frac{\mathrm{d}}{\mathrm{d}t}A(t)\right]+\left[\frac{\mathrm{d}}{\mathrm{d}t}A(t)\right]A(t)$$

确实不同于(5.4).更一般地,对应于指数 k 为任何正整数的幂函数

$f(u)=u^k$,容易用归纳法证明

$$\frac{\mathrm{d}}{\mathrm{d}t}A^k=\dot{A}A^{k-1}+A\frac{\mathrm{d}}{\mathrm{d}t}A^{k-1}$$
$$=\dot{A}A^{k-1}+A\dot{A}A^{k-2}+\cdots+A^{k-1}\dot{A}. \tag{5.5}$$

9.5.8 定理 设 p 是任一多项式, A 是可微 $n\times n$ 矩阵值函数; A 关于实变量 t 的导数记作 $\dot{A}$, 则

(1) 如果对于 t 的一特殊值矩阵 A 与 $\dot{A}$ 可交换,那么对于 t 的该值成立形如(5.4)的链法则:

$$\frac{\mathrm{d}}{\mathrm{d}t}p(A)=p'(A)\dot{A}. \tag{5.6}$$

(2) 即使 A 与 $\dot{A}$ 不可交换,链法则对于迹仍能成立:

$$\frac{\mathrm{d}}{\mathrm{d}t}\left[\mathrm{tr}p(A)\right]=\mathrm{tr}\left[p'(A)\dot{A}\right]. \tag{5.7}$$

证 假设 A 与 $\dot{A}$ 可交换,则(5.5)可以写成

$$\frac{\mathrm{d}}{\mathrm{d}t}A^k=kA^{k-1}\dot{A}.$$

这是公式(5.6)当 $p(u)=u^k$ 的情形;因为所有多项式是幂的线性组合,应用求导的线性性质,便推出(5.6)对所有多项式成立.

对于不可交换的 A 与 $\dot{A}$,考虑(5.5)的迹.依 **1.5.16**,迹具有交换性:

$$\mathrm{tr}\left(A^i\dot{A}A^{k-i-1}\right)=\mathrm{tr}\left(A^{k-i-1}A^i\dot{A}\right)=\mathrm{tr}\left(A^{k-1}\dot{A}\right).$$

由此推出

$$\mathrm{tr}\left(\frac{\mathrm{d}}{\mathrm{d}t}A^k\right)=k\,\mathrm{tr}\left(A^{k-1}\dot{A}\right).$$

又依 **9.5.5** 的(4),迹与求导可以交换,因此(5.7)对 $p(u)=u^k$ 的情形成立.再应用求导的线性性质,便可以将(5.7)推广于任何多项式. □

9.5.9 定理 设 M 是 k 元多重线性函数，$x^{(1)},\cdots,x^{(k)}$ 是可微向量值函数，则 $M\left(x^{(1)},\cdots,x^{(k)}\right)$ 可微，而且

$$\frac{\mathrm{d}}{\mathrm{d}t}M\left(x^{(1)},\cdots,x^{(k)}\right)=M\left(\frac{\mathrm{d}}{\mathrm{d}t}x^{(1)},x^{(2)}\cdots,x^{(k)}\right)+\cdots+M\left(x^{(1)},\cdots,x^{(k-1)},\frac{d}{dt}x^{(k)}\right). \tag{5.8}$$

证 因为 M 是多重线性的，

$$\begin{aligned}
&M\left(x^{(1)}(t+h),\cdots,x^{(k)}(t+h)\right)-M\left(x^{(1)}(t),\cdots,x^{(k)}(t)\right)\\
&=M\left(x^{(1)}(t+h)-x^{(1)}(t),x^{(2)}(t+h),\cdots,x^{(k)}(t+h)\right)\\
&\quad+M\left(x^{(1)}(t),x^{(2)}(t+h)-x^{(2)}(t),x^{(3)}(t+h),\cdots,x^{(k)}(t+h)\right)\\
&\quad+\cdots+M\left(x^{(1)}(t),\cdots,x^{(k-1)}(t),x^{(k)}(t+h)-x^{(k)}(t)\right),
\end{aligned}$$

等式两边除以 h 且让 $h\to 0$ 即得(5.8). □

(5.8)的最重要应用之一是关于 **1.5.7** 定义的行列式所确定的函数 D：

$$\frac{\mathrm{d}}{\mathrm{d}t}D\left(x^{(1)},\cdots,x^{(n)}\right)=D\left(\frac{\mathrm{d}}{\mathrm{d}t}x^{(1)},x^{(2)}\cdots,x^{(n)}\right)+\cdots+D\left(x^{(1)},\cdots,x^{(n-1)},\frac{\mathrm{d}}{\mathrm{d}t}x^{(n)}\right). \tag{5.9}$$

9.5.10 定理 设 A 是可微 $n\times n$ 矩阵值函数，则对于使 $A(t)$ 可逆的 t 成立

$$\frac{\mathrm{d}}{\mathrm{d}t}\left[\log(\det A)\right]=\operatorname{tr}\left(A^{-1}\frac{\mathrm{d}}{\mathrm{d}t}A\right). \tag{5.10}$$

证 先考虑可微按列分块矩阵 $X=\left[x^{(1)}\quad\cdots\quad x^{(n)}\right]$，其中

$$x^{(i)}=\left(x_{1i},\cdots,x_{ni}\right)^{\mathrm{T}},\quad i=1,\cdots,n.$$

并设 X 满足 $X(t_0)=I$，亦即 $x^{(i)}(t_0)=e_i$ 是第 i 分量为1其余分量

为 0 的单位列向量,$i=1,\cdots,n$.此时,(5.9)右边的各行列式在 $t=t_0$ 处的值为

$$D\left(\left.\frac{\mathrm{d}}{\mathrm{d}t}x^{(1)}(t)\right|_{t=t_0},e_2,\cdots,e_n\right)=\dot{x}_{11}(t_0),$$

$$D\left(e_1,\left.\frac{\mathrm{d}}{\mathrm{d}t}x^{(2)}(t)\right|_{t=t_0},e_3,\cdots,e_n\right)=\dot{x}_{22}(t_0),$$

$$\cdots$$

$$D\left(e_1,\cdots,e_{n-1},\left.\frac{\mathrm{d}}{\mathrm{d}t}x^{(n)}(t)\right|_{t=t_0}\right)=\dot{x}_{nn}(t_0),$$

因此

$$\left.\frac{\mathrm{d}}{\mathrm{d}t}\left[\det X(t)\right]\right|_{t=t_0}=\mathrm{tr}\dot{X}(t_0), \tag{5.11}$$

现设 $t=t_0$ 处 A 可逆,并定义

$$X(t)\equiv[A(t_0)]^{-1}A(t),$$

则

$$A(t)=A(t_0)X(t), \tag{5.12}$$

且 $X(t_0)=I$ 可以应用(5.11).对(5.12)取行列式,

$$\det A(t)=\det A(t_0)\det X(t),$$

并将其及(5.12)代入(5.11),得

$$\left[\det A(t_0)\right]^{-1}\left.\frac{\mathrm{d}}{\mathrm{d}t}\left[\det A(t)\right]\right|_{t=t_0}=\mathrm{tr}\left[\left(A(t_0)\right)^{-1}\dot{A}(t_0)\right].$$

然后将左端写成

$$\left.\frac{\mathrm{d}}{\mathrm{d}t}\left[\log\left(\det A(t)\right)\right]\right|_{t=t_0},$$

即得(5.10). □

此结果的重要性在于它建立了行列式与迹的一种联系.

9.5.11 定义 设 A 是任一 $n\times n$ 矩阵,定义**矩阵指数函数** e^A 为

$$e^A=\sum_{k=0}^{\infty}\frac{1}{k!}A^k. \tag{5.13}$$

这一定义是对纯量指数函数 Taylor 级数形式的推广.收敛性的证明与纯量的情形是相同的,问题在于证明部分和之差趋向于零.用 $e_m(A)$ 表示(5.13)的第 m 部分和:

$$e_m(A)=\sum_{k=0}^{m}\frac{1}{k!}A^k, \tag{5.14}$$

则

$$e_m(A)-e_l(A)=\sum_{k=l+1}^{m}\frac{1}{k!}A^k. \tag{5.15}$$

利用矩阵范数的次加性和次乘性(见 **1.8.28**),得

$$\|e_m(A)-e_l(A)\|\le\sum_{k=l+1}^{m}\frac{1}{k!}\|A\|^k. \tag{5.16}$$

这样便回到纯量的情形,因此可以估计(5.16)的右端,并推出随 l 和 m 趋向无穷,(5.15)的右端趋向于零,而且对范数 $\|A\|$ 小于事先任意指定常数的所有矩阵 A 来说是一致的.

矩阵指数函数具有纯量指数函数的若干(不是全部)性质.

9.5.12 定理 (1) 若 A 与 B 是可交换的 $n\times n$ 矩阵,则

$$e^{A+B}=e^Ae^B=e^Be^A.$$

(2) 若 A 与 B 是不可交换的 $n\times n$ 矩阵,则一般地

$$e^{A+B}\ne e^Ae^B\ne e^Be^A.$$

(3) 若 A 是定义在一个区间上的矩阵值函数,且对自变量 t 可微,则 e^A 也对 t 可微.

(4) 若对 t 的一特殊值, $A(t)$ 与 $\dot{A}(t)$ 可交换,则

$$\frac{\mathrm{d}}{\mathrm{d}t}e^{A}=\mathrm{e}^{A}\dot{A}=\dot{A}\mathrm{e}^{A}.$$

(5) 若 A 是斜自伴的, $A^{*}=-A$,则 e^{A} 是正交的.

证 利用 A 与 B 可交换,有展开式

$$(A+B)^{k}=\sum_{i=1}^{k}\binom{k}{i}A^{i}B^{k-i},$$

于是从定义 **9.5.11** 推出(1).

对于(2),举如下例子:

$$A=\begin{bmatrix}0&1\\0&0\end{bmatrix},\qquad B=\begin{bmatrix}0&0\\1&0\end{bmatrix},$$

直接验证知 $AB\neq BA$.由于 $A^{2}=0,B^{2}=0$,依 **9.5.11**,

$$\mathrm{e}^{A}=I+A=\begin{bmatrix}1&1\\0&1\end{bmatrix},\ \mathrm{e}^{B}=I+B=\begin{bmatrix}1&0\\1&1\end{bmatrix}.$$

于是

$$\mathrm{e}^{A}\mathrm{e}^{B}=\begin{bmatrix}2&1\\1&1\end{bmatrix},\qquad \mathrm{e}^{B}\mathrm{e}^{A}=\begin{bmatrix}1&1\\1&2\end{bmatrix}.$$

另一方面,注意到

$$A+B=\begin{bmatrix}0&1\\1&0\end{bmatrix};$$

$$(A+B)^{2k}=I=\begin{bmatrix}1&0\\0&1\end{bmatrix},\ (A+B)^{2k+1}=\begin{bmatrix}0&1\\1&0\end{bmatrix}.$$

因此

$$\mathrm{e}^{A+B}=\sum_{k=0}^{\infty}\frac{1}{k!}(A+B)^{k}$$

$$=\begin{bmatrix}1&0\\0&1\end{bmatrix}\sum_{k=0}^{\infty}\frac{1}{(2k)!}+\begin{bmatrix}0&1\\1&0\end{bmatrix}\sum_{k=0}^{\infty}\frac{1}{(2k+1)!}$$

$$=\frac{e^1+e^{-1}}{2}\begin{bmatrix}1 & 0\\0 & 1\end{bmatrix}+\frac{e^1-e^{-1}}{2}\begin{bmatrix}0 & 1\\1 & 0\end{bmatrix}$$

$$=(\cosh 1)\begin{bmatrix}1 & 0\\0 & 1\end{bmatrix}+(\sinh 1)\begin{bmatrix}0 & 1\\1 & 0\end{bmatrix},$$

所以$e^{A+B}\neq e^A e^B\neq e^B e^A$.

(3)的证明依赖如下矩阵情形模拟微分学的一个重要性质:设$\{G_m\}$是定义在一个区间上的可微矩阵值函数序列,满足条件

① $\{G_m\}$一致收敛于极限函数G,

② 导函数序列$\{\dot{G}_m\}$一致收敛于极限函数F,

则G是可微的,而且$\dot{G}=F$.这一结论的证明留作练习.

将这一结论应用于级数(5.13)的部分和序列:

$$G_m(t)\equiv e_m(A(t)),\quad m=0,1,2,\cdots$$

这里记号$e_m(A)$由(5.14)定义.因为在一个区间上A可微,所以G_m和$\dot{G}_m$有意义,$\{G_m\}$一致收敛于e^A,$\{\dot{G}_m\}$也一致收敛.于是,e^A可微,而且$\{\dot{G}_m\}$一致收敛于$\frac{d}{dt}e^A$.

(4)的关于$\frac{d}{dt}e^A$的显式公式可以通过对级数(5.13)的逐项求导而得到.

为了证明(5),对无穷级数(5.13)逐项取伴随:

$$(e^A)^*=\sum_{k=0}^{\infty}\frac{1}{k!}(A^k)^*=\sum_{k=0}^{\infty}\frac{1}{k!}(A^*)^k=e^{A^*}=e^{-A}.$$

然后利用(1),有

$$(e^A)^*e^A=e^{-A}e^A=e^0=I,$$

这表明e^A是正交矩阵. □

9.5.13 定理 特征值是连续地依赖于矩阵的.其含义如下:如果$\{A_m\}$

是收敛的 $n\times n$ 矩阵序列，$\lim\limits_{m\to\infty}A_m=A$，那么 A_m 的特征值集收敛于 A 的特征值集. 换句话说，对每个 $\varepsilon>0$，存在正整数 k，使得当 $m>k$ 时，A_m 的所有特征值属于 A 的各特征值以其自身为中心以 ε 为半径的圆盘之并.

证 A_m 的特征值是特征多项式 $p_m(\lambda)\equiv\det(\lambda I-A_m)$ 的根. 因为 A_m 趋向于 A，亦即 A_m 的所有元素趋向于 A 的相应元素；推出 p_m 的系数趋向 A 的特征多项式 p 的系数. 又因为多项式的根连续地依赖于系数，因此定理结论成立. □

从这一定理出发，转向讨论矩阵单重特征值依赖于矩阵的方式.

9.5.14 定理 设 A 是实变量 t 的可微 $n\times n$ 矩阵值函数，$A(0)$ 有单重特征值 λ_0，其含义为 λ_0 是 $A(0)$ 的特征多项式的单根，则对于足够小的 t，$A(t)$ 有一个特征值 $\lambda(t)$，它对 t 可微，而且 $\lambda(0)=\lambda_0$.

证 $A(t)$ 的特征多项式

$$p(\lambda,t)\equiv\det(\lambda I-A(t))$$

是 λ 的 n 次多项式，其系数是 t 的可微函数. λ_0 是 $A(0)$ 的单重特征值的假设蕴涵着

$$p(\lambda_0,0)=0,\quad \left.\frac{\partial}{\partial\lambda}p(\lambda,0)\right|_{\lambda=\lambda_0}\neq 0$$

根据隐函数定理，在如此条件下，方程

$$p(\lambda,t)=0$$

有解 $\lambda=\lambda(t)$，而且这个解在 $t=0$ 的一个邻域中对 t 可微. □

在此定理的条件下，可以选取属于特征值 $\lambda(t)$ 的特征向量是对 t 可微的. 说"可以选取"只是因为特征向量总可以差一个不可微的纯量函数因子. 为了证明这样的结论需要如下引理.

9.5.15 引理 设 A 是 $n\times n$ 矩阵，p 是其特征多项式，λ_0 是 p 的单根，则 $A-\lambda_0 I$ 至少有一个 $(n-1)\times(n-1)$ 主子矩阵的行列式不为零.

证 只须从 A 减 $\lambda_0 I$ 就可把问题归结成特征值为零，因此不妨直接假设 $\lambda_0=0$.

现在 0 是 p 的单根，成立

$$p(0)=0,\quad \left(\frac{\mathrm{d}p}{\mathrm{d}\lambda}\right)(0)\neq 0.$$

为了计算 p 的导数，以 $a^{(1)},\cdots,a^{(n)}$ 表示 A 的列，$e_1,\cdots,e_n$ 是单位向量，则

$$\lambda I-A=\begin{bmatrix}\lambda e_1-a^{(1)} & \lambda e_2-a^{(2)} & \cdots & \lambda e_n-a^{(n)}\end{bmatrix}.$$

依(5.9)，

$$\begin{aligned}\left(\frac{\mathrm{d}p}{\mathrm{d}\lambda}\right)(0)&=\frac{\mathrm{d}}{\mathrm{d}\lambda}\det(\lambda I-A)\bigg|_{\lambda=0}\\&=\det\begin{bmatrix}e_1 & -a^{(2)} & \cdots & -a^{(n)}\end{bmatrix}\\&\quad+\det\begin{bmatrix}-a^{(1)} & e_2 & \cdots & -a^{(n)}\end{bmatrix}\\&\quad+\cdots+\det\begin{bmatrix}-a^{(1)} & \cdots & -a^{(n-1)} & e_n\end{bmatrix}.\end{aligned}$$

显然，等式右端是 n 个 $(n-1)\times(n-1)$ 主子矩阵的行列式之和的 $(-1)^{n-1}$ 倍.因为 $\left(\frac{\mathrm{d}p}{\mathrm{d}\lambda}\right)(0)\neq 0$，所以这些主子矩阵的行列式至少有一个是非零的. □

9.5.16 定理 设 A 是定理 **9.5.14** 中描述的矩阵值函数，$\lambda(t)$ 是那里描述的 $A(t)$ 的特征值，则可选取 $A(t)$ 属于 $\lambda(t)$ 的特征向量 $x(t)$ 是对 t 可微的.

证 首先证明，对于引理 **9.5.15** 中的 A，如果 $A-\lambda_0 I$ 的第 i 个主子矩阵(即 $A-\lambda_0 I$ 删去第 i 行和第 i 列后的剩余部分)的行列式

不为零,那么 A 属于 λ_0 的特征向量 $x^{(0)}$ 的第 i 个分量必不为零.如若不然,用 $\hat{x}^{(0)}$ 表示 $x^{(0)}$ 删去第 i 个分量而得的向量,用 A_{ii} 表示 A 的第 i 个主子矩阵,则 $\hat{x}^{(0)}$ 满足

$$(A_{ii}-\lambda_0 I)\hat{x}^{(0)}=0. \tag{5.17}$$

因 $\det(A_{ii}-\lambda_0 I)\neq 0$, $A_{ii}-\lambda_0 I$ 可逆,从(5.17)得 $\hat{x}^{(0)}=0$.于是从 $x^{(0)}$ 的第 i 个分量等于零的假设得 $x^{(0)}=0$.这是一个矛盾,因为特征向量不能为零.

这样,作为规范化 $x^{(0)}$ 的一种方式,可取 $x^{(0)}$ 的第 i 个分量为 1,它的其余分量即 $\hat{x}^{(0)}$ 的分量,从 $(A-\lambda_0 I)x^{(0)}=0$ 推知满足非齐次方程

$$(A_{ii}-\lambda_0 I)\hat{x}^{(0)}=-\hat{a}^{(i)}, \tag{5.18}$$

其中 $\hat{a}^{(i)}$ 是 A 的第 i 列删去第 i 个分量而得的向量.所以

$$\hat{x}^{(0)}=-(A_{ii}-\lambda_0 I)^{-1}\hat{a}^{(i)}. \tag{5.19}$$

现在回到本定理的假设,矩阵 $A(0)$ 和特征值 λ_0 满足引理 **9.5.15** 的条件.从而,因矩阵 $A_{ii}(0)-\lambda_0 I$ 可逆,且因 $A(t)$ 连续地依赖于 t,由 **9.5.13** 推出,对于足够小的 t, $A_{ii}(t)-\lambda(t)I$ 可逆,故可以取相应 $\lambda(t)$ 的特征向量 $x(t)$ 的第 i 个分量为 1,并由公式(5.19)来确定 $x(t)$ 的其余分量:

$$\hat{x}(t)=-(A_{ii}(t)-\lambda(t)I)^{-1}(\hat{a}(t))^{(i)}, \tag{5.20}$$

其中 $\hat{x}(t)$ 是 $x(t)$ 删去第 i 个分量而得的向量,$(\hat{a}(t))^{(i)}$ 是 $A(t)$ 的第 i 列 $(a(t))^{(i)}$ 删去第 i 个分量而得的向量.

因为(5.20)右端对 t 可微,所以 $\hat{x}(t)$ 对 t 可微.于是,$x(t)$ 对 t 可微. □

9.5.17 引理 设 A 是 $n\times n$ 矩阵,p 是其特征多项式,λ_0 是 p 的 k 重根,则 $A-\lambda_0 I$ 的零空间至多为 k 维.

证 不失一般性,考虑$\lambda_0=0$为p的k重根,于是有

$$p(0)=\frac{\mathrm{d}p}{\mathrm{d}\lambda}(0)=\cdots=\frac{\mathrm{d}^{k-1}p}{\mathrm{d}\lambda^{k-1}}(0)=0,\ \frac{\mathrm{d}^{k}p}{\mathrm{d}\lambda^{k}}(0)\neq 0.$$

作为引理 **9.5.15** 的延伸,对$\det(\lambda I-A)$求导k次,可将p在0的k阶导数表示成$(n-k)\times(n-k)$主子矩阵的行列式之和.因为k阶导数不为0,推出这些行列式至少有一个是非零的,比如说是A删去第1,…,第k的行和列而得的主子矩阵$\hat{A}\equiv A(\{k+1,\cdots,n\})$,
$\det\hat{A}\neq 0$.

这样,可以证明A的零空间$\mathscr{N}_A$不包含前k个分量为零的非零向量.如若不然,设$x^{(0)}$是前k个分量为零的非零向量而且$x^{(0)}\in\mathscr{N}_A$;用$\hat{x}^{(0)}$表示$x^{(0)}$删去前k个分量而得的向量.因为$Ax^{(0)}=0$,经缩短的向量$\hat{x}^{(0)}$满足

$$\hat{A}\hat{x}^{(0)}=0. \tag{5.21}$$

又因为$\det\hat{A}\neq 0$,$\hat{A}$可逆;所以从(5.21)推出$\hat{x}^{(0)}=0$,并因此$x^{(0)}=0$.这是一个矛盾.

现在即可推出$\dim\mathscr{N}_A\le k$.事实上,假若$\dim\mathscr{N}_A>k$,那么依**1.3.4** 的(1),$\mathscr{N}_A$中存在非零向量x满足k个线性条件

$$x_1=0,\ \cdots,\ x_k=0,$$

亦即x的前k个分量为零.然而,刚刚证明了$\mathscr{N}_A$中不包含如此非零向量. □

这一引理是 **9.5.15** 的扩展,是一个有用的结论.

9.5.18 定理 设A是$n\times n$矩阵,p是其特征多项式,λ_0是p的k重根,则A属于特征值λ_0的广义特征向量空间的维数等于k.

证 依 **1.6.15**,广义特征向量空间是$(A-\lambda_0 I)^d$的零空间,其中d是特征值λ_0的指数.不妨设$\lambda_0=0$,从而只须证明A^d的零空间

的维数等于 k.

由于

$$\tau I - A^d = \prod_{j=0}^{d-1}\left(\tau^{1/d} I - \omega^j A\right),$$

其中 ω 是本原 d 次单位根.取行列式并利用行列式的乘法性质,得 A^d 的特征多项式 p_d 通过 A 的特征多项式 p 的如下表示式:

$$\begin{aligned} p_d(\tau) = \det\left(\tau I - A^d\right) &= \prod_{j=0}^{d-1}\det\left(\tau^{1/d} I - \omega^j A\right) \\ &= \pm\prod_{j=0}^{d-1}\det\left(\omega^{-j}\tau^{1/d} I - A\right) \\ &= \pm\prod_{j=0}^{d-1} p\left(\omega^{-j}\tau^{1/d}\right). \end{aligned} \tag{5.22}$$

注意到 $\lambda_0 = 0$ 是 p 的 k 重根, p 可以写成

$$p(\tau) = \hat{p}(\tau)\tau^k,$$

其中 $\hat{p}$ 是 $n-k$ 次多项式, $\hat{p}(0) \neq 0$.因此,从(5.22)推出 0 也是 p_d 的 k 重根.这样,依 **9.5.17**, A^d 的零空间的维数 $\leq k$;下面证明等号成立.

记 p 的根为 $\lambda_1, \cdots, \lambda_m$,重数分别为 $k_1, \cdots, k_m$,

$$k_1 + \cdots + k_m = n. \tag{5.23}$$

用 $\mathscr{N}_i$ 表示 A 属于特征值 λ_i 的广义特征向量空间,依 **1.6.13**,每个向量可以分解为广义特征向量之和,换言之,有

$$\mathbf{C}^n = \mathscr{N}_1 \oplus \cdots \oplus \mathscr{N}_m,$$

因而

$$n = \dim \mathscr{N}_1 + \cdots + \dim \mathscr{N}_m. \tag{5.24}$$

由于 $\mathscr{N}_i$ 即 $\left(A - \lambda_i I\right)^{d_i}$ 的零空间, d_i 是特征值 λ_i 的指数;上面已经证明

$$\dim \mathscr{N}_i \le k_i, \quad i = 1, \cdots, m. \tag{5.25}$$

联合(5.23),(5.24)和(5.25),得

$$n \le k_1 + \cdots + k_m = n.$$

由此推出(5.25)中的所有不等式成立等号. □

关于多重特征值情形,一般矩阵难免出现广义特征向量,以致更难对其依赖矩阵的方式作出分析.为此,多讨论自伴矩阵,原因是它们没有广义特征向量.自伴矩阵多重特征值分析在第 3 章中作了介绍.

9.6 λ-矩阵

9.6.1 定义 设 $A(\lambda) = [a_{ij}(\lambda)]$ 是 $m \times n$ 矩阵,其每个元素 $a_{ij}(\lambda)$ 均为 λ 的多项式,则称 $A(\lambda)$ 为 λ**-矩阵**.所有 $a_{ij}(\lambda)$ 的最大次数,称为 $A(\lambda)$ 的**次数**,记作 $\deg A(\lambda)$. $A(\lambda)$ 中为非零 λ 多项式的子式的最大阶数,称为 $A(\lambda)$ 的**秩**.

设 $A(\lambda)$ 是 $m \times n$ 的 λ-矩阵.如果 $\deg A(\lambda) = k$,那么 $A(\lambda)$ 可以写成

$$A(\lambda) = A_k \lambda^k + A_{k-1} \lambda^{k-1} + \cdots + A_1 \lambda + A_0, \tag{6.1}$$

其中 $A_k, \cdots, A_0 \in \mathbb{C}^{m \times n}$.

特别,当 $A_k = I$ 时,称 $A(\lambda)$ 为**首一**(monic)λ**-矩阵**.

自然,每个 $A \in \mathbb{C}^{m \times n}$ 可视为 λ-矩阵的特殊情形.

λ-矩阵无非是以一元多项式为元素的矩阵,只是传统上以 λ 为参变元而已.这是一类重要的函数矩阵.

9.6.2 定义 设$A(\lambda)$是$n\times n$的λ-矩阵.如果$\det A(\lambda)\neq 0$(即不恒等于零),则称$A(\lambda)$是**正则**的.如果存在$n\times n$的λ-矩阵$B(\lambda)$,使得

$$A(\lambda)B(\lambda)=B(\lambda)A(\lambda)=I,$$

则称$A(\lambda)$是**可逆**的;换言之,$A(\lambda)$可逆是指$B(\lambda)=(A(\lambda))^{-1}$存在而且也是$\lambda$-矩阵.

9.6.3 定理 λ-矩阵$A(\lambda)$可逆的充分必要条件是$\det A(\lambda)$等于非零常数.

证 若$\det A(\lambda)=c\neq 0$,则$A(\lambda)$的逆矩阵的元素等于$A(\lambda)$的$n-1$阶子式除以c,从而是λ的多项式.因此,$(A(\lambda))^{-1}$是λ-矩阵.

反之,若$A(\lambda)$可逆,则从等式$A(\lambda)B(\lambda)=I$得

$$\det A(\lambda)\det B(\lambda)=1.$$

这是说多项式$\det A(\lambda)$和多项式$\det B(\lambda)$之积是非零常数.显然,仅当两者皆为非零常数时才可能如此. □

行列式等于非零常数的λ-矩阵也称为**幺模λ-矩阵**.因此,幺模λ-矩阵就是可逆λ-矩阵,两者等价.

9.6.4 定义 设$A(\lambda)$和$B(\lambda)$是$n\times n$的λ-矩阵,且$\deg A(\lambda)=l$和$\deg B(\lambda)=m$,

$$A(\lambda)=\sum_{i=0}^{l}\lambda^i A_i,\qquad B(\lambda)=\sum_{i=0}^{m}\lambda^i B_i,$$

其中$A_0,\cdots,A_l,B_0,\cdots,B_m\in\mathbb{C}^{n\times n}$.

$A(\lambda)$和$B(\lambda)$可按通常的矩阵加法和乘法计算$A(\lambda)+B(\lambda)$和$A(\lambda)B(\lambda)$.

现在考虑$A(\lambda)$和$B(\lambda)$的除法.如果$B(\lambda)$的首项矩阵B_m可逆,而且存在$n\times n$的λ-矩阵$Q(\lambda)$和$R(\lambda)$,$\deg R(\lambda)<m$, 使得

$$A(\lambda)=Q(\lambda)B(\lambda)+R(\lambda),$$

则称这是 $B(\lambda)$**右除** $A(\lambda)$,$Q(\lambda)$称为$B(\lambda)$除$A(\lambda)$的**右商**,$R(\lambda)$称为$B(\lambda)$除$A(\lambda)$的**右余**.特别,$R(\lambda)\equiv 0$时,称$Q(\lambda)$为$B(\lambda)$除$A(\lambda)$的**右因子**.

类似地, 如果$B(\lambda)$的首项矩阵B_m可逆,而且存在$n\times n$的λ-矩阵$\hat{Q}(\lambda)$和$\hat{R}(\lambda)$,$\deg\hat{R}(\lambda)<m$,使得

$$A(\lambda)=B(\lambda)\hat{Q}(\lambda)+\hat{R}(\lambda),$$

则称这是$B(\lambda)$**左除**$A(\lambda)$,$\hat{Q}(\lambda)$称为$B(\lambda)$除$A(\lambda)$的**左商**,$\hat{R}(\lambda)$称为$B(\lambda)$除$A(\lambda)$的**左余**.特别,$\hat{R}(\lambda)\equiv 0$时,称$\hat{Q}(\lambda)$为$B(\lambda)$除$A(\lambda)$的**左因子**.

9.6.5 定理 设$A(\lambda)=\sum\limits_{i=0}^{l}\lambda^i A_i$和$B(\lambda)=\sum\limits_{i=0}^{m}\lambda^i B_i$是$n\times n$的$\lambda$-矩阵,

$$\deg A(\lambda)=l,\quad \deg B(\lambda)=m,$$

且$\det B_m\neq 0$,则存在唯一的$B(\lambda)$除$A(\lambda)$的右商和右余,以及左商和左余.

证 只证$B(\lambda)$右除$A(\lambda)$情形,左除情形相仿.如果$l<m$,可以并只能取$Q(\lambda)=0$和$R(\lambda)=A(\lambda)$,结论成立.

现设$l\geq m$.首先,考虑"除以"$B(\lambda)$的首项$B_m\lambda^m$.容易看出λ-矩阵$A_lB_m^{-1}\lambda^{l-m}B(\lambda)$的最高次项正好是$A_l\lambda^l$.因此

$$A(\lambda)=A_lB_m^{-1}\lambda^{l-m}B(\lambda)+A^{(1)}(\lambda),$$

其中$A^{(1)}(\lambda)\equiv A(\lambda)-A_lB_m^{-1}\lambda^{l-m}B(\lambda)$,将其表示为

$$A^{(1)}(\lambda)=A_{l_1}^{(1)}\lambda^{l_1}+\cdots+A_1^{(1)}\lambda+A_0^{(1)},\quad A_{l_1}^{(1)}\neq 0,$$

显然,$l_1\leq l-1$.

如果仍有$l_1\geq m$,则将$A(\lambda)$代之以$A^{(1)}(\lambda)$,重复以上过程,

有

$$A^{(1)}(\lambda)=A_{l_1}^{(1)}B_m^{-1}\lambda^{l_1-m}B(\lambda)+A^{(2)}(\lambda),$$

其中

$$A^{(2)}(\lambda)=A_{l_2}^{(2)}\lambda^{l_2}+\cdots+A_1^{(2)}\lambda+A_0^{(2)},\quad A_{l_2}^{(1)}\neq 0,$$

而且 $l_2\leq l_1-1$.

如此,可以构造一个 λ-矩阵序列:

$$A^{(s-1)}(\lambda)=A_{l_{s-1}}^{(s-1)}B_m^{-1}\lambda^{l_{s-1}-m}B(\lambda)+A^{(s)}(\lambda),\ s=1,\cdots,r,$$

其中

$$A^{(0)}(\lambda)\equiv A(\lambda).$$

并且

$$\deg A^{(r-1)}\equiv l_{r-1}\geq m,\quad \deg A^{(r)}\equiv l_r<m.$$

联合这些关系式,得

$$A(\lambda)=Q(\lambda)B(\lambda)+R(\lambda),$$

其中

$$Q(\lambda)\equiv A_l B_m^{-1}\lambda^{l-m}+A_{l_1}^{(1)}B_m^{-1}\lambda^{l_1-m}+\cdots+A_{l_{r-1}}^{(r-1)}B_m^{-1}\lambda^{l_{r-1}-m}$$

和

$$R(\lambda)\equiv A^{(r)}(\lambda)$$

分别是 $B(\lambda)$ 除 $A(\lambda)$ 的右商和右余.

现在假设还存在 $\hat{Q}(\lambda)$ 和 $\hat{R}(\lambda)$ 满足

$$A(\lambda)=\hat{Q}(\lambda)B(\lambda)+\hat{R}(\lambda),\quad \deg\hat{R}(\lambda)<m,$$

则成立

$$\left(Q(\lambda)-\hat{Q}(\lambda)\right)B(\lambda)=\hat{R}(\lambda)-R(\lambda).$$

如果 $\hat{Q}(\lambda)\neq Q(\lambda)$,那么上面等式左端的 λ-矩阵的次数至少为 m,而右端的 λ-矩阵的次数必小于 m,矛盾.于是

$$\hat{Q}(\lambda)=Q(\lambda).$$

并因此 $\hat{R}(\lambda)=R(\lambda)$. □

9.6.6 定理 设 $A(\lambda)=\sum_{i=0}^{l}\lambda^i A_i$ 是 $n\times n$ 的 λ-矩阵，$B\in\mathbb{C}^{n\times n}$，则 $\lambda I-B$ 除 $A(\lambda)$的右余是

$$A(B)\equiv A_l B^l + A_{l-1}B^{l-1}+\cdots+A_1 B+A_0;$$

左余是

$$\hat{A}(B)\equiv B^l A_l + B^{l-1}A_{l-1}+\cdots+BA_1+A_0.$$

证 容易验证成立因式分解

$$\lambda^j I-B^j=\left(\lambda^{j-1}I+\lambda^{j-2}B+\cdots+\lambda B^{j-2}+B^{j-1}\right)(\lambda I-B),$$

在等式两端左乘 A_j，然后对 $j=1,\cdots,l$ 的所得等式求和.和式的右端形如

$$C(\lambda)(\lambda I-B),$$

其中 $C(\lambda)$是 λ-矩阵;左端为

$$\begin{aligned}\sum_{j=1}^{l}A_j\lambda^j-\sum_{j=1}^{l}A_jB^j&=\sum_{j=0}^{l}A_j\lambda^j-\sum_{j=0}^{l}A_jB^j\\&=A(\lambda)-A(B),\end{aligned}$$

因此

$$A(\lambda)=C(\lambda)(\lambda I-B)+A(B).$$

由此,并依 **9.6.5**, $\lambda I-B$ 除 $A(\lambda)$的右余是 $A(B)$.

左余的结果可以通过对调因式分解中的因子,右乘以 A_j,然后求和而推得. □

9.6.7 定义 两个 $m\times n$ 的 λ-矩阵 $A(\lambda)$和 $B(\lambda)$称为**等价**的,或者说是**连接等价变换**的,如果存在 $m\times m$ 幺模 λ-矩阵 $P(\lambda)$和 $n\times n$ 幺模 λ-矩阵 $Q(\lambda)$,使得

$$P(\lambda)A(\lambda)Q(\lambda)=B(\lambda). \tag{6.2}$$

9.6.8 定义 设 $P(\lambda)=\left[p_{ij}(\lambda)\right]$是 $n\times n$ 非奇异 λ-矩阵.有如下三种简单类型矩阵:

(1) $\begin{cases} p_{ii}(\lambda)=1, \quad i=1,\cdots,n, \quad i\neq p,q, \\ p_{pq}(\lambda)=p_{qp}(\lambda)=1, \\ p_{ij}(\lambda)=0, \quad \text{其它}. \end{cases}$

此即置换矩阵,形如

$$\begin{bmatrix} 1 & & & & & & & & & \\ & \ddots & & & & & & & & \\ & & 1 & & & & & & & \\ & & & 0 & \cdots & \cdots & \cdots & 1 & & \\ & & & \vdots & 1 & & & \vdots & & \\ & & & \vdots & & \ddots & & \vdots & & \\ & & & \vdots & & & 1 & \vdots & & \\ & & & 1 & \cdots & \cdots & \cdots & 0 & & \\ & & & & & & & & 1 & \\ & & & & & & & & & \ddots & \\ & & & & & & & & & & 1 \end{bmatrix} \begin{matrix} \\ \\ \\ p \\ \\ \\ \\ q \\ \\ \\ \\ \end{matrix}$$

(2) $\begin{cases} p_{ii}(\lambda)=1, \quad i=1,\cdots,n, \\ p_{pq}(\lambda)=f(\lambda), \quad p\neq q, \\ p_{ij}(\lambda)=0, \quad \text{其它}. \end{cases}$

形如

$$\begin{bmatrix} 1 & & & & \\ & \ddots & \cdots & f(\lambda) & \\ & & \ddots & \vdots & \\ 0 & & & \ddots & \\ & & & & 1 \end{bmatrix} \begin{matrix} \\ p \\ \\ q \\ \\ \end{matrix} \quad \text{或} \quad \begin{bmatrix} 1 & & & & \\ & \ddots & & 0 & \\ & \vdots & \ddots & & \\ & f(\lambda) & \cdots & \ddots & \\ & & & & 1 \end{bmatrix} \begin{matrix} \\ p \\ \\ q \\ \\ \end{matrix}$$

$$(3)\ \begin{cases} p_{ii}(\lambda)=1, & i=1,\cdots,n, \quad i\neq p, \\ p_{pp}(\lambda)=c\neq 0, & \\ p_{ij}(\lambda)=0, & \text{其它}. \end{cases}$$

形如

$$\begin{bmatrix} 1 & & & & & & \\ & \ddots & & & & \text{\Huge 0} & \\ & & 1 & & & & \\ & & & c & & & \\ & & & & 1 & & \\ & \text{\Huge 0} & & & & \ddots & \\ & & & & & & 1 \end{bmatrix} p$$

这些矩阵称为**初等变换矩阵**.

用初等变换矩阵左(右)乘λ-矩阵$A(\lambda)$分别起如下作用:

(1) 交换第p和第q行(列).

(2) 第q行(列)的$f(\lambda)$倍加至第p行(列).

(3) 第p行(列)乘以非零常数c.

对$A(\lambda)$的这些变换称为**初等变换**.

显然,三类初等变换的行列式分别等于-1,1和c.

9.6.9 定理 设$A(\lambda)=[a_{ij}(\lambda)]$是$n\times n$ λ-矩阵,$\operatorname{rank}A(\lambda)=r$,则可以通过初等变换将$A(\lambda)$约化为等价的如下形式的$\lambda$-矩阵:

$$D(\lambda)=\begin{bmatrix} \hat{D}(\lambda) & 0 \\ 0 & 0 \end{bmatrix}, \quad \hat{D}(\lambda)=\operatorname{diag}(a_1(\lambda),\cdots,a_r(\lambda)), \tag{6.3}$$

其中 $a_1(\lambda),\cdots,a_r(\lambda)$ 均是首项系数为1的 λ 的多项式,而且 $a_i(\lambda)$ 整除 $a_{i+1}(\lambda), i=1,\cdots,r-1$.

证 在下面,利用一系列初等变换来产生形如(6.3)的矩阵;而且无论进行到任何阶段,总把现行变换中的矩阵的 (i,j)-元素叫做 $a_{ij}(\lambda)$.

设 $a_{ij}(\lambda)$ 是 $A(\lambda)$ 中次数最低的(非零)元素之一;通过第1和第 i 行交换及第1和第 j 列交换,该元素便成为元素 $a_{11}(\lambda)$.这样,可得如下表示:

$$\begin{aligned} a_{i1}(\lambda)&=a_{11}(\lambda)q_{i1}(\lambda)+r_{i1}(\lambda), \quad i=2,\cdots,m, \\ a_{1j}(\lambda)&=a_{11}(\lambda)q_{1j}(\lambda)+r_{1j}(\lambda), \quad j=2,\cdots,n, \end{aligned} \tag{6.4}$$

其中 $r_{21}(\lambda),\cdots,r_{m1}(\lambda)$ 和 $r_{12}(\lambda),\cdots,r_{1n}(\lambda)$ 是余项,它们的次数低于 $a_{11}(\lambda)$ 的次数.如果它们不全为零,假设 $r_{k1}(\lambda)$ 不为零;那么,将第 k 行减去 $q_{k1}(\lambda)$ 倍的第1行,然后交换第1和第 k 行.于是,元素 $a_{11}(\lambda)$ 改变为 $r_{k1}(\lambda)$,它不为零,而且其次数低于前此在这个位置上的元素的次数.

继续如此做法,元素 $a_{11}(\lambda)$ 总是非零而且其次数必定减少.

最终,或者 $a_{11}(\lambda)$ 变换成与 λ 无关,或者在更早些阶段,现行 $a_{11}(\lambda)$ 必然整除所有现行 $a_{21}(\lambda),\cdots,a_{m1}(\lambda)$ 和 $a_{12}(\lambda),\cdots,a_{1n}(\lambda)$. 这时,通过2至 m 行分别减去适当倍数的第1行,2至 n 列分别减去适当倍数的第1列,矩阵便可约化成如下形式:

$$\begin{bmatrix} a_{11}(\lambda) & 0 \\ 0 & A_2(\lambda) \end{bmatrix}. \tag{6.5}$$

至此,看 $a_{11}(\lambda)$ 是否可以整除 $A_2(\lambda)$ 的所有元素.如若不然,假设对于某元素 $a_{ij}(\lambda)$,

$$a_{ij}(\lambda)=a_{11}(\lambda)q_{ij}(\lambda)+r_{ij}(\lambda), \quad r_{ij}(\lambda)\neq 0,$$

那么,将第 i 行加至第1行,然后施行前面的步骤,重新得到(6.5)形式,但非零的 $a_{11}(\lambda)$ 具有更低的次数.继续如此做法,或者 $a_{11}(\lambda)$ 变换

成与 λ 无关,或者在更早些阶段,必定达到形如(6.5)且 $a_{11}(\lambda)$ 能整除 $A_2(\lambda)$ 的所有元素的情形.

如果 $A_2(\lambda)$ 不恒等于零,可以把对 $A(\lambda)$ 的做法用来处理 $A_2(\lambda)$,得到

$$\left[\begin{array}{cc|c} a_{11}(\lambda) & 0 & \\ 0 & a_{22}(\lambda) & 0 \\ \hline \multicolumn{2}{c|}{0} & A_3(\lambda) \end{array}\right], \tag{6.6}$$

其中 $a_{11}(\lambda)$ 整除 $a_{22}(\lambda)$, $a_{22}(\lambda)$ 整除 $A_3(\lambda)$ 的所有元素.

继续这样做下去,矩阵约化为形如

$$\left[\begin{array}{ccc|c} a_{11}(\lambda) & & 0 & 0 \\ & \ddots & & \vdots \\ 0 & & a_{ss}(\lambda) & 0 \\ \hline 0 & \cdots & 0 & 0 \end{array}\right] \tag{6.7}$$

其中每个 $a_{ii}(\lambda)$ 整除其后继者;当 s 等于 m 或 n,或者 $A_{s+1}(\lambda)$ 为零矩阵时处理终止.

由于 λ-矩阵在用非奇异矩阵相乘下不会改变秩,而且根据假设, $\operatorname{rank}A(\lambda)=r$,故 $s=r$.因为每个 $a_{ii}(\lambda)$ 是非零多项式,乘其所在行以适当的常数,可使其最高次幂的系数等于1,从而每个 $a_{ii}(\lambda)$ 成为首项系数为1的多项式,简记之 $a_i(\lambda)$, $a_i(\lambda)$ 整除 $a_{i+1}(\lambda)$.这样,最后得到了(6.3)中的 $D(\lambda)$.鉴于 $D(\lambda)$ 是由一系列初等变换而获得的,因此有

$$P(\lambda)A(\lambda)Q(\lambda)=D(\lambda), \tag{6.8}$$

其中 $P(\lambda)$ 和 $Q(\lambda)$ 分别是若干初等变换矩阵之积,必为非奇异的,而且行列式之值与 λ 无关. □

这是关于等价 λ-矩阵的基本定理.(6.3)中的 $D(\lambda)$ 称为等价 λ-矩阵 $A(\lambda)$ 的 **Smith 标准形**.

9.6.10 定理 λ-矩阵的秩在等价变换下是不变的.

证 设λ-矩阵$A(\lambda)$和$B(\lambda)$等价,

$$B(\lambda)=P(\lambda)A(\lambda)Q(\lambda),$$

则$B(\lambda)$的j阶子式$b(\lambda)$可以通过$A(\lambda)$的若干j阶子式$a_s(\lambda)$表示:

$$b(\lambda)=\sum_s p_s(\lambda)a_s(\lambda)q_s(\lambda), \qquad (6.9)$$

其中$p_s(\lambda)$和$q_s(\lambda)$分别是$P(\lambda)$和$Q(\lambda)$的某些j阶子式.现在假设$B(\lambda)$的秩为r,$b(\lambda)$是$B(\lambda)$的r阶非零子式,于是从(6.9)推出,至少存在$A(\lambda)$的一个r阶子式$a_s(\lambda)$是非零多项式,这表明

$$\operatorname{rank}B(\lambda)\le\operatorname{rank}A(\lambda).$$

注意到依**9.6.7**和**9.6.3**,$P(\lambda)$和$Q(\lambda)$是可逆的.因而从

$$A(\lambda)=[P(\lambda)]^{-1}B(\lambda)[Q(\lambda)]^{-1}, \qquad (6.10)$$

又可推出

$$\operatorname{rank}A(\lambda)\le\operatorname{rank}B(\lambda). \qquad \square$$

9.6.11 定义 设λ-矩阵$A(\lambda)$的秩为r,用$d_j(\lambda)$表示$A(\lambda)$所有j阶子式的**首一最大公因子**,$j=1,\cdots,r$,并规定$d_0(\lambda)\equiv 1$.(首一指首项系数为1.)

显然,任何$j\ge 2$阶的子式可以表示为$j-1$阶子式的线性组合,$d_{j-1}(\lambda)$必为$d_j(\lambda)$的因子,故序列

$$d_0(\lambda),d_1(\lambda),\cdots,d_r(\lambda) \qquad (6.11)$$

中,$d_j(\lambda)$可被$d_{j-1}(\lambda)$整除,$j=1,\cdots,r$.因此,商

$$i_j(\lambda)\equiv\frac{d_j(\lambda)}{d_{j-1}(\lambda)}, \quad j=1,\cdots,r \qquad (6.12)$$

都是多项式.

注意,

$$d_j(\lambda)=i_1(\lambda)\cdots i_j(\lambda),\quad j=1,\cdots,r, \tag{6.13}$$

而且$i_j(\lambda)$可被$i_{j-1}(\lambda)$整除, $j=2,\cdots,r$.

$i_1(\lambda),\cdots,i_r(\lambda)$称为$A(\lambda)$的**不变多项式**.这一术语基于如下定理.

9.6.12 定理 设λ-矩阵$A(\lambda)$和$B(\lambda)$的秩为r,而且是等价的,$d_j(\lambda)$和$\delta_j(\lambda)$分别是$A(\lambda)$和$B(\lambda)$所有j阶子式的首一最大公因子, $j=1,\cdots,r$,则

$$d_j(\lambda)=\delta_j(\lambda),\quad j=1,\cdots,r. \tag{6.14}$$

证 沿用**9.6.10**证明中的记号.

从(6.9)看出,$A(\lambda)$的各个j阶子式的任何公因子是$b(\lambda)$的因子,故$\delta_j(\lambda)$可被$d_j(\lambda)$整除, $j=1,\cdots,r$.

仍然,从(6.10)看出,$d_j(\lambda)$又可被$\delta_j(\lambda)$整除, $j=1,\cdots,r$.

再注意到所有多项式$d_j(\lambda)$和$\delta_j(\lambda)$的首项系数均为1,推出(6.14)成立. □

9.6.13 定理 设$A(\lambda)$是秩为r的λ-矩阵,(6.3)中的$D(\lambda)$是$A(\lambda)$的Smith标准形,则

$$a_j(\lambda)=i_j(\lambda),\quad j=1,\cdots,r, \tag{6.15}$$

其中$i_1(\lambda),\cdots,i_r(\lambda)$是$A(\lambda)$的不变多项式.

证 $D(\lambda)$的子式的最大公因子序列为

$$d_j(\lambda)=a_1(\lambda)\cdots a_j(\lambda),\quad j=1,\cdots,r.$$

依**9.6.12**,它们也是$A(\lambda)$的子式的最大公因子序列.因此

$$i_j(\lambda)=\frac{d_j(\lambda)}{d_{j-1}(\lambda)}=a_j(\lambda),\quad j=1,\cdots,r. \qquad \square$$

根据这一结论,**9.6.9**可以改述如下.

设 $A(\lambda)$ 是秩为 r 的 $m\times n$ 的 λ-矩阵, $i_1(\lambda),\cdots,i_r(\lambda)$ 是 $A(\lambda)$ 的不变多项式，则 $A(\lambda)$ 等价于

$$D(\lambda)=\begin{bmatrix}\hat{D}(\lambda) & 0\\ 0 & 0\end{bmatrix},\quad \hat{D}(\lambda)=\operatorname{diag}(i_1(\lambda),\cdots,i_r(\lambda)). \tag{6.16}$$

9.6.14 推论 两个 λ-矩阵等价的充分必要条件是有相同的不变多项式.

证明留作练习.

9.6.15 定理 设 $A,B\in\mathbb{C}^{n\times n}$, A 和 B 相似的充分必要条件是 λ-矩阵 $\lambda I-A$ 和 $\lambda I-B$ 等价.

或者,利用 **9.6.14**, A 和 B 相似的充分必要条件是 λ-矩阵 $\lambda I-A$ 和 $\lambda I-B$ 有相同的不变多项式.

证 如果 A 和 B 相似,那么存在非奇异矩阵 $S\in\mathbb{C}^{n\times n}$,使得 $A=SBS^{-1}$,从而

$$\lambda I-A=S(\lambda I-B)S^{-1}.$$

这表明 $\lambda I-A$ 和 $\lambda I-B$ 是等价的.

反之,如果 $\lambda I-A$ 和 $\lambda I-B$ 有相同的不变多项式,那么它们是等价的.这就是说,存在幺模 λ-矩阵 $P(\lambda)$ 和 $Q(\lambda)$,使得

$$P(\lambda)(\lambda I-A)Q(\lambda)=\lambda I-B. \tag{6.17}$$

于是

$$M(\lambda)(\lambda I-B)=(\lambda I-A)Q(\lambda), \tag{6.18}$$

其中

$$M(\lambda)\equiv[P(\lambda)]^{-1}$$

也是一个 λ-矩阵.

现在用 $\lambda I-A$ 左除 $M(\lambda)$,用 $\lambda I-B$ 右除 $Q(\lambda)$,有

$$M(\lambda)=(\lambda I-A)S(\lambda)+M_0 \tag{6.19}$$

和

$$Q(\lambda)=R(\lambda)(\lambda I-B)+Q_0, \tag{6.20}$$

其中$M_0,Q_0\in\mathbb{C}^{n\times n}$.将(6.19)和(6.20)代入(6.18),得

$$((\lambda I-A)S(\lambda)+M_0)(\lambda I-B)=(\lambda I-A)(R(\lambda)(\lambda I-B)+Q_0).$$

由此,

$$(\lambda I-A)(S(\lambda)-R(\lambda))(\lambda I-B)=(\lambda I-A)Q_0-M_0(\lambda I-B).$$

注意到等式右端λ-矩阵的次数至多为1,而如果$S(\lambda)-R(\lambda)$不为零,则左端λ-矩阵次数至少为2.推出

$$S(\lambda)\equiv R(\lambda).$$

并因此

$$M_0(\lambda I-B)=(\lambda I-A)Q_0. \tag{6.21}$$

这表明

$$M_0=Q_0,\ \ M_0B=AQ_0.$$

从而有

$$M_0B=AM_0. \tag{6.22}$$

至此,只须论证M_0非奇异便可完结定理的证明.

设用$\lambda I-B$左除$P(\lambda)$得

$$P(\lambda)=(\lambda I-B)U(\lambda)+P_0,$$

其中P_0与λ无关.因此,利用(6.19),(6.20)和(6.21)有

$$\begin{aligned}I&=M(\lambda)P(\lambda)\\&=((\lambda I-A)S(\lambda)+M_0)((\lambda I-B)U(\lambda)+P_0)\\&=(\lambda I-A)S(\lambda)(\lambda I-B)U(\lambda)+(\lambda I-A)Q_0U(\lambda)\\&\quad+(\lambda I-A)S(\lambda)P_0+M_0P_0\\&=(\lambda I-A)(Q(\lambda)U(\lambda)+S(\lambda)P_0)+M_0P_0.\end{aligned}$$

从而推出

$$Q(\lambda)U(\lambda)+S(\lambda)P_0=0,\quad M_0P_0=I.$$

于是,$\det M_0\neq 0$,(6.22)蕴涵A和B相似. □

9.6.16 定理 设 $A\in\mathbb{C}^{n\times n}$, $\lambda I-A$ 的非零次不变多项式是

$$i_s(\lambda), i_{s+1}(\lambda), \cdots, i_n(\lambda),$$

则 A 相似于块对角矩阵

$$C_1=\operatorname{diag}\left(C_{i_s}, C_{i_{s+1}}, \cdots, C_{i_n}\right), \tag{6.23}$$

其中 C_{i_k} 表示相伴不变多项式 $i_k(\lambda)$ 的友矩阵, $s\le k\le n$.

证 首先,注意到 λ-矩阵 $\lambda I-A$ 的秩为 n 以及关于其不变多项式的假设,存在幺模 λ-矩阵 $P(\lambda)$ 和 $Q(\lambda)$,使得

$$P(\lambda)(\lambda I-A)Q(\lambda)=\operatorname{diag}\left(1,\cdots,1,i_s(\lambda),\cdots,i_n(\lambda)\right). \tag{6.24}$$

对这一等式取行列式,左端因 $\det P(\lambda)$ 和 $\det Q(\lambda)$ 是非零常数,等于数乘 A 的特征多项式 $\det(\lambda I-A)$,次数为 n;而右端等于乘积 $i_s(\lambda)\cdots i_n(\lambda)$.由此可知 $C_1\in\mathbb{C}^{n\times n}$.

其次,注意到相伴首一多项式

$$p(\lambda)=\lambda^m+a_{m-1}\lambda^{m-1}+\cdots+a_1\lambda+a_0$$

的友矩阵形如

$$C_p=\begin{bmatrix} 0 & 1 & & \text{\Large 0} \\ & \ddots & \ddots & \\ \text{\Large 0} & & 0 & 1 \\ -a_0 & -a_1 & \cdots & -a_{m-1} \end{bmatrix}.$$

容易计算, $\lambda I-C_p$ 的不变多项式为

$$i_1(\lambda)=\cdots=i_{m-1}(\lambda)=1, \quad i_m(\lambda)=p(\lambda).$$

从而 C_p 等价于 $m\times m$ 对角矩阵 $\operatorname{diag}(1,\cdots,1,p(\lambda))$.

因此,对于(6.23)中的 C_{i_k}, $s\le k\le n$, λ-矩阵 $\lambda I-C_{i_k}$ 等价于对角矩阵

$$D_k(\lambda)\equiv\operatorname{diag}\left(1,\cdots,1,i_k(\lambda)\right).$$

从而推出 $\lambda I-C_1$ 等价于

$$D(\lambda)\equiv\operatorname{diag}\left(D_s(\lambda),\cdots,D_n(\lambda)\right).$$

具体地,存在幺模 λ-矩阵 $\hat{P}(\lambda)$和 $\hat{Q}(\lambda)$,使得

$$\lambda I - C_1 = \hat{P}(\lambda)D(\lambda)\hat{Q}(\lambda). \tag{6.25}$$

显然,通过初等变换矩阵可将 $D(\lambda)$ 变换为(6.24)中右端的形式,即

$$\Delta(\lambda) \equiv \operatorname{diag}(1,\cdots,1,i_s(\lambda),\cdots,i_n(\lambda)),$$

故 $D(\lambda)$ 和 $\Delta(\lambda)$ 等价.于是,从(6.25), $\lambda I - C_1$ 和 $\Delta(\lambda)$ 等价;这说明

$$1,\cdots,1,i_s(\lambda),\cdots,i_n(\lambda)$$

是 $\lambda I - C_1$ 和 $\lambda I - A$ 两者的不变多项式.从而,依 **9.6.15**, A 和 C_1 相似. □

对比 **5.2.4**, C_1 就是 A 的 Frobenius 标准形,然而这里的刻画更为精细些,论证也更加具体些. C_1 也称为 A 的**第一自然标准形**(first natural normal form).

9.6.17 定义 设 $A(\lambda)$是 $n \times n$ 的 λ-矩阵, $\operatorname{rank}A(\lambda) = r$,其不变多项式为

$$i_1(\lambda),\cdots,i_r(\lambda).$$

且设 $d_r(\lambda)$是 $A(\lambda)$所有 r 阶子式的首一最大公因子,在C上,

$$d_r(\lambda) = \prod_{j=1}^{s}(\lambda - \lambda_j)^{m_j}, \tag{6.26}$$

其中

$$m_j \geq 1, \quad j = 1,\cdots,s,$$

而 $\lambda_1,\cdots,\lambda_s$ 是相异的数,它们称为 $A(\lambda)$的**本征根**(latent root).

注意到

$$i_1(\lambda)\cdots i_r(\lambda) = d_r(\lambda),$$

以及 $i_{j-1}(\lambda)$是 $i_j(\lambda)$的因子, $j = 2,\cdots,r$.推出存在一组整数

$$\alpha_{jk} \geq 0; \qquad 1 \leq j \leq r,\ 1 \leq k \leq s,$$

使得

$$\begin{aligned}
i_1(\lambda) &= (\lambda-\lambda_1)^{\alpha_{11}} (\lambda-\lambda_2)^{\alpha_{12}} \cdots (\lambda-\lambda_s)^{\alpha_{1s}}, \\
i_2(\lambda) &= (\lambda-\lambda_1)^{\alpha_{21}} (\lambda-\lambda_2)^{\alpha_{22}} \cdots (\lambda-\lambda_s)^{\alpha_{2s}}, \\
&\cdots \\
i_r(\lambda) &= (\lambda-\lambda_1)^{\alpha_{r1}} (\lambda-\lambda_2)^{\alpha_{r2}} \cdots (\lambda-\lambda_s)^{\alpha_{rs}},
\end{aligned} \tag{6.27}$$

其中

$$0 \leq \alpha_{1k} \leq \alpha_{2k} \leq \cdots \leq \alpha_{rk} \leq m_k,$$

$$\sum_{j=1}^{r} \alpha_{jk} = m_k, \quad k = 1, \cdots, s. \tag{6.28}$$

出现在因子分解(6.27)中的每个 $\alpha_{jk} > 0$ 的因子 $(\lambda - \lambda_k)^{\alpha_{jk}}$ 称为 $A(\lambda)$的(关于 λ_k 的)**初等因子**.

通常,$\alpha_{jk} = 1$ 的初等因子称为**线性初等因子**;否则称为**非线性初等因子**.

9.6.18 定理 两个 λ-矩阵等价的充分必要条件是两者有相同的初等因子.

证 依 **9.6.14**,立即推出初等因子在 λ-矩阵的等价变换下是不变的. □

9.6.19 定理 设 $A, B \in \mathbb{C}^{n\times n}$, A 和 B 相似的充分必要条件是 λ-矩阵 $\lambda I - A$ 和 $\lambda I - B$ 有相同的初等因子.

证 直接从 **9.6.15** 和 **9.6.18** 推出. □

9.6.20 定理 设 $A(\lambda)$和 $B(\lambda)$是 λ-方阵,则块对角矩阵

$$C(\lambda) = \begin{bmatrix} A(\lambda) & 0 \\ 0 & B(\lambda) \end{bmatrix}$$

的初等因子集合等于$A(\lambda)$的初等因子集合和$B(\lambda)$的初等因子集合的并集.

证 设$D_1(\lambda)$和$D_2(\lambda)$分别是$A(\lambda)$和$B(\lambda)$的 Smith 标准形,则显然存在幺模λ-矩阵$E(\lambda)$和$F(\lambda)$,使得

$$C(\lambda)=E(\lambda)\begin{bmatrix} D_1(\lambda) & 0 \\ 0 & D_2(\lambda) \end{bmatrix}F(\lambda)$$

现在设

$$(\lambda-\lambda_0)^{\alpha_1},\cdots,(\lambda-\lambda_0)^{\alpha_p}$$

和

$$(\lambda-\lambda_0)^{\beta_1},\cdots,(\lambda-\lambda_0)^{\beta_q}$$

分别是$D_1(\lambda)$和$D_2(\lambda)$关于同一$\lambda_0\in\mathbb{C}$的初等因子.并设按非减顺序排列指数集:

$$\{\gamma_1,\cdots,\gamma_{p+q}\}\equiv\{\alpha_1,\cdots,\alpha_p,\beta_1,\cdots,\beta_q\},$$

其中

$$0\le\gamma_1\le\cdots\le\gamma_{p+q}.$$

依 **9.6.11**,在$\operatorname{diag}(D_1(\lambda),D_2(\lambda))$的 Smith 标准形

$$D=\operatorname{diag}(i_1(\lambda),\cdots,i_r(\lambda),0,\cdots,0)$$

中的不变多项式,

$i_r(\lambda)$可被$(\lambda-\lambda_0)^{\gamma_{p+q}}$整除,不能被$(\lambda-\lambda_0)^{\gamma_{p+q}+1}$整除;

$i_{r-1}(\lambda)$可被$(\lambda-\lambda_0)^{\gamma_{p+q-1}}$整除,不能被$(\lambda-\lambda_0)^{\gamma_{p+q-1}+1}$整除;

...

由此推出

$$\begin{bmatrix} D_1(\lambda) & 0 \\ 0 & D_2(\lambda) \end{bmatrix},$$

——从而$C(\lambda)$——关于λ_0的初等因子正好是

$$(\lambda-\lambda_0)^{\gamma_1},\cdots,(\lambda-\lambda_0)^{\gamma_{p+q}}.$$

□

9.6.21 第二自然标准形定理 设$A \in \mathbb{C}^{n\times n}$,用

$$l_1(\lambda),\cdots,l_s(\lambda)$$

表示λ-矩阵$\lambda I - A$的初等因子,则A相似于块对角矩阵

$$C_2 = \mathrm{diag}(C_{l_1},\cdots,C_{l_s}), \tag{6.29}$$

其中C_{l_k}是相伴多项式$l_k(\lambda)$的友矩阵,$1 \le k \le s$.

证 矩阵$\lambda I - C_{l_k}$仅有的非常数不变多项式是$l_k(\lambda)$,又因作为$\lambda I - A$的初等因子的$l_k(\lambda)$是λ的线性多项式的幂,故$l_k(\lambda)$也是$\lambda I - C_{l_k}$的唯一的初等因子.

应用 **9.6.20**,$\lambda I - C_2$和$\lambda I - A$有相同的初等因子.从而,依 **9.6.19**,A和C_2相似. □

应用这一定理可以极其简洁地证明关于 Jordan 标准形的结论 **5.1.2**.

9.6.22 Jordan 标准形定理 设$A \in \mathbb{C}^{n\times n}$,$\lambda I - A$的初等因子为

$$l_k(\lambda) \equiv (\lambda - \lambda_k)^{n_k}, \quad k = 1,\cdots,s, \ \sum_{k=1}^{s} n_k = n, \tag{6.30}$$

则A相似于矩阵

$$J = \mathrm{diag}(J_{n_1},\cdots,J_{n_s}) \in \mathbb{C}^{n\times n}, \tag{6.31}$$

其中J_{n_k}是相应于$l_k(\lambda)$的$n_k \times n_k$ Jordan 块,

$$J_{n_k} = \begin{bmatrix} \lambda_k & 1 & & 0 \\ & \ddots & \ddots & \\ & & \ddots & 1 \\ 0 & & & \lambda_k \end{bmatrix}, \quad k = 1,\cdots,s. \tag{6.32}$$

证 $\lambda I - J_{n_k}$仅有初等因子$l_k(\lambda)$.

另一方面,由 **9.6.21** 的证明得知,$\lambda I - C_{l_k}$同样仅有初等因子$l_k(\lambda)$,C_{l_k}是相伴$l_k(\lambda)$的友矩阵.

因此,矩阵 J_{n_k} 和 C_{l_k} 相似,$k=1,\cdots,s$.从而,依 **9.6.20**,矩阵 J 和(6.29)中的 C_2 相似.

现在应用 **9.6.21** 即得 A 和 J 相似. □

9.7 有理矩阵

9.7.1 定义 设 $A(\lambda)=[a_{ij}(\lambda)]$ 是 $m\times n$ 矩阵,其每个元素 $a_{ij}(\lambda)$ 均为 λ 的有理式,则称 $A(\lambda)$ 为 λ 的**有理矩阵**(rational matrix).

λ 的有理式是一个分式,其分子和分母是 λ 的多项式.分子的次数小于分母的次数的有理式称为**真有理式**.如果每个 $a_{ij}(\lambda)$ 均为真有理式,则称 $A(\lambda)$ 为**严格真有理矩阵**(strictly proper rational matrix).

显然,若 $A(\lambda)$ 是严格真有理矩阵,则当 $\lambda\to\infty$ 时,$A(\lambda)\to 0$.

9.7.2 定理 设 $A(\lambda)=[a_{ij}(\lambda)]$ 是 $m\times n$ 有理矩阵,则存在 $m\times m$ 幺模 λ-矩阵 $P(\lambda)$ 和 $n\times n$ 幺模 λ-矩阵 $Q(\lambda)$,使得

$$P(\lambda)A(\lambda)Q(\lambda)=D(\lambda)\equiv\begin{bmatrix}\hat{D}(\lambda) & 0\\ 0 & 0\end{bmatrix}, \tag{7.1}$$

其中

(1) $\hat{D}(\lambda)\equiv\operatorname{diag}(a_1(\lambda),\cdots,a_r(\lambda))$,其中

$$a_i(\lambda)\equiv\varphi_i(\lambda)/\psi_i(\lambda),\quad i=1,\cdots,r,$$

每对 $\varphi_i(\lambda)$ 和 $\psi_i(\lambda)$ 是首项系数为1的互素的多项式;

(2) $\varphi_i(\lambda)$ 是 $\varphi_{i+1}(\lambda)$ 的因子,$i=1,\cdots,r-1$;

(3) $\psi_{i+1}(\lambda)$ 是 $\psi_i(\lambda)$ 的因子,$i=1,\cdots,r-1$.

证 用 $g(\lambda)$ 表示 $A(\lambda)$ 所有元素分母的首项系数为1的最小公倍式,则 $g(\lambda)A(\lambda)$ 是 λ-矩阵.设

$$\operatorname{rank}[g(\lambda)A(\lambda)] = r .$$

依 **9.6.9**, $g(\lambda)A(\lambda)$ 可约化为 Smith 标准形

$$P(\lambda)[g(\lambda)A(\lambda)]Q(\lambda) = D_s(\lambda) \equiv \begin{bmatrix} \hat{D}_s(\lambda) & 0 \\ 0 & 0 \end{bmatrix}, \qquad (7.2)$$

其中

$$\hat{D}_s(\lambda) \equiv \operatorname{diag}(b_1(\lambda), \cdots, b_r(\lambda)),$$

$b_1(\lambda), \cdots, b_r(\lambda)$ 都是首项系数为1的多项式,而且 $b_i(\lambda)$ 是 $b_{i+1}(\lambda)$ 的因子, $i = 1, \cdots, r-1$.

现在对等式(7.2)除以 $g(\lambda)$,记

$$D(\lambda) \equiv \frac{1}{g(\lambda)} D_s(\lambda),$$

$$\hat{D}(\lambda) \equiv \frac{1}{g(\lambda)} \hat{D}_s(\lambda) = \operatorname{diag}\left(\frac{b_1(\lambda)}{g(\lambda)}, \cdots, \frac{b_r(\lambda)}{g(\lambda)} \right).$$

再对 $i = 1, \cdots, r$,将 $b_i(\lambda)/g(\lambda)$ 消去 $b_i(\lambda)$ 和 $g(\lambda)$ 的公因子,并记其为

$$a_i(\lambda) \equiv \varphi_i(\lambda)/\psi_i(\lambda).$$

这样,从(7.2)推得(7.1).

从推证过程看出条件(1)成立.注意到 $b_i(\lambda)$ 是 $b_{i+1}(\lambda)$ 的因子, $i = 1, \cdots, r-1$,条件(2)和(3)也是成立的. □

(7.1)中的 $D(\lambda)$ 称为有理矩阵 $A(\lambda)$ 的 **Smith-McMillan(标准)形**.

$\psi_1(\lambda), \cdots, \psi_r(\lambda)$ 的次数之和称为 $A(\lambda)$ 的 **McMillan 次数**.

$\varphi_1(\lambda), \cdots, \varphi_r(\lambda)$ 的根称为 $A(\lambda)$ 的**零点**.

$\psi_1(\lambda), \cdots, \psi_r(\lambda)$ 的根称为 $A(\lambda)$ 的**极点**.

例 矩阵

$$A(\lambda)=\begin{bmatrix}\dfrac{1}{(\lambda+3)^2} & \dfrac{1}{(\lambda+3)(\lambda+4)}\\ \dfrac{-6}{(\lambda+3)(\lambda+4)^2} & \dfrac{\lambda-1}{(\lambda+4)^2}\end{bmatrix}\tag{7.3}$$

是严格真有理矩阵,其所有元素的最小公分母是

$$g(\lambda)=(\lambda+3)^2(\lambda+4)^2.$$

这样,

$$g(\lambda)A(\lambda)=\begin{bmatrix}(\lambda+4)^2 & (\lambda+3)(\lambda+4)\\ -6(\lambda+3) & (\lambda-1)(\lambda+3)^2\end{bmatrix},\tag{7.4}$$

且 $\det[g(\lambda)A(\lambda)]\neq 0$, $\mathrm{rank}[g(\lambda)A(\lambda)]=2$.

因(7.4)中四个元素没有公因式,故 $g(\lambda)A(\lambda)$ 的 Smith 标准形 $D_s(\lambda)$ 的对角元素是

$b_1(\lambda)=1$,

$b_2(\lambda)=\det[g(\lambda)A(\lambda)]=(\lambda+1)(\lambda+2)(\lambda+3)^2(\lambda+4)$,

所以,(7.3)中的 $A(\lambda)$ 的 Smith-McMillan 形是

$$D(\lambda)=\frac{1}{g(\lambda)}D_s(\lambda)=\begin{bmatrix}\dfrac{1}{g(\lambda)} & 0\\ 0 & \dfrac{b_2(\lambda)}{g(\lambda)}\end{bmatrix}$$

$$=\begin{pmatrix}\dfrac{1}{(\lambda+3)^2(\lambda+4)^2} & 0\\ 0 & \dfrac{(\lambda+1)(\lambda+2)}{(\lambda+4)}\end{pmatrix}.$$

在这里,

$\varphi_1(\lambda)=1,\qquad \psi_1(\lambda)=(\lambda+3)^2(\lambda+4)^2;$

$\varphi_2(\lambda)=(\lambda+1)(\lambda+2),\qquad \psi_2(\lambda)=\lambda+4.$

$A(\lambda)$的 McMillan 次数为5,不同于$g(\lambda)$的次数.

$A(\lambda)$的零点是$-1,-2$.

$A(\lambda)$的极点是-3(2次),-4(3次).

注意,虽然(7.3)中的$A(\lambda)$是严格真有理矩阵,但是它的 Smith-McMillan 形却不是.

9.7.3 定义 设$A\in\mathbb{C}^{n\times n},B\in\mathbb{C}^{n\times l},C\in\mathbb{C}^{m\times n},D\in\mathbb{C}^{m\times l}$,方程组

$$\begin{cases}\lambda x = Ax + Bu, \\ y = Cx + Du,\end{cases} \quad x\in\mathbb{C}^n, u\in\mathbb{C}^l, y\in\mathbb{C}^m, \tag{7.5}$$

其中λ是复参数.

容易推出

$$y = G(\lambda)u, \tag{7.6}$$

其中

$$G(\lambda)\equiv C(\lambda I - A)^{-1}B + D. \tag{7.7}$$

$G(\lambda)$是λ的$m\times l$有理矩阵,称为关于方程组(7.5)的**转移(或传递)函数矩阵**(transfer function matrix).

在线性控制系统理论中,方程组(7.5)起着基础的作用,在那里,x是状态向量,u是控制向量,y是输出向量.

9.7.4 定义 对于给定的严格真有理矩阵$G(\lambda)$,确定一组矩阵A,B,C,使得

$$C(\lambda I - A)^{-1}B = G(\lambda), \tag{7.8}$$

称A,B,C为$G(\lambda)$的一个**实现**(realization).

如此实现是不唯一的,其中具有最小阶数的A的实现(对应于状态变量个数最少)称为**极小**(minimal)**实现**.

注意,在$G(\lambda)$是严格真有理矩阵的条件下,必有$D=0$.因此,确

定 $G(\lambda)$ 的实现是确定转移函数矩阵的反问题.

容易证明(留作练习):

若矩阵 A,B,C 是 $G(\lambda)$ 的一个实现,则对于任何与 A 同阶的非奇异矩阵 T,矩阵 TAT^{-1},TB,CT^{-1} 也 $G(\lambda)$ 的实现.

9.7.5 定理 $A\in\mathbb{C}^{n\times n},B\in\mathbb{C}^{n\times l},C\in\mathbb{C}^{m\times n}$ 为 $G(\lambda)$ 的极小实现的充分必要条件是:

(1) 矩阵对 (A,B) 为**可控制**的,即**可控制矩阵**(见 **3.14.15**)

$$\mathscr{C}\equiv\begin{bmatrix}B & AB & A^2B & \cdots & A^{n-1}B\end{bmatrix} \tag{7.9}$$

的秩 rank$\mathscr{C}=n$;

(2) 矩阵对 (A,C) 为**可观测**的(observable),即**可观测矩阵**

$$\begin{bmatrix}C\\ CA\\ CA^2\\ \vdots\\ CA^{n-1}\end{bmatrix} \tag{7.10}$$

的秩等于 n.

定理证明从略.

例 设 A 是 $n\times n$ 友矩阵

$$A=\begin{bmatrix}0 & 1 & 0 & \cdots & 0\\ \vdots & \ddots & \ddots & \ddots & \vdots\\ \vdots & & \ddots & \ddots & 0\\ 0 & \cdots & \cdots & 0 & 1\\ -a_n & \cdots & \cdots & -a_2 & -a_1\end{bmatrix},$$

其特征多项式

$$p(\lambda) \equiv \det(\lambda I - A) = \lambda^n + a_1\lambda^{n-1} + \cdots + a_{n-1}\lambda + a_n .$$

设

$$v(\lambda) = \left(1, \lambda, \lambda^2, \cdots, \lambda^{n-1}\right)^{\mathrm{T}} .$$

不难推证：$(\lambda I - A)^{-1}$ 的最后一列是 $v(\lambda)/p(\lambda)$；从而纯量转移函数

$$g(\lambda) \equiv \left(c_0\lambda^{n-1} + c_1\lambda^{n-2} + \cdots + c_{n-2}\lambda + c_{n-1}\right)/p(\lambda) \quad (7.11)$$

的实现是

$$A,\ b = e_n \equiv (0, \cdots, 0, 1)^{\mathrm{T}},\ c = \left(c_{n-1}, \cdots, c_1, c_0\right). \quad (7.12)$$

亦即成立

$$c(\lambda I - A)^{-1} b = g(\lambda).$$

而且容易验证这一实现的矩阵对 (A, b) 是可控制的.

因此，依 **9.7.5**，(7.11)中 $g(\lambda)$ 的实现(7.12)为极小实现的充分必要条件是可观测矩阵

$$\begin{bmatrix} c \\ cA \\ cA^2 \\ \vdots \\ cA^{n-1} \end{bmatrix}$$

为非奇异矩阵.然而，依 **6.8.3**，这个 $n \times n$ 可观测矩阵正好是关于多项式 $p(\lambda)$ 和 $c_0\lambda^{n-1} + c_1\lambda^{n-2} + \cdots + c_{n-2}\lambda + c_{n-1}$ 的结式矩阵. 于是，可以换一种说法：

(7.12)为极小实现的充分必要条件是(7.11)中分子和分母两个多项式互素.

9.7.6 定义 设 $A(\lambda)$ 是 $m \times n$ 有理矩阵，用 $g_i(\lambda)$ 表示 $A(\lambda)$ 的第 i 行中元素的最小公分母，用 $\tilde{g}_j(\lambda)$ 表示 $A(\lambda)$ 的第 j 列中元素的最小

公分母,并记

$$\begin{aligned} M(\lambda) &\equiv \operatorname{diag}(g_1(\lambda),\cdots,g_m(\lambda)), \\ \tilde{M}(\lambda) &\equiv \operatorname{diag}(\tilde{g}_1(\lambda),\cdots,\tilde{g}_n(\lambda)), \end{aligned} \tag{7.13}$$

则 $A(\lambda)$ 可表示为

$$A(\lambda)=(M(\lambda))^{-1}N(\lambda) \tag{7.14}$$

和

$$A(\lambda)=\tilde{N}(\lambda)(\tilde{M}(\lambda))^{-1}, \tag{7.15}$$

其中 $N(\lambda)$ 和 $\tilde{N}(\lambda)$ 是 $m\times n$ 的 λ-矩阵.

表示式(7.14)称为 $A(\lambda)$ 的**左矩阵分式描述**(left matrix fraction description)或简称**左 m.f.d.**, $N(\lambda)$ 为其**分子**, $M(\lambda)$ 为其**分母**.

表示式(7.15)称为 $A(\lambda)$ 的**右矩阵分式描述**(right matrix fraction description)或简称**右 m.f.d.**, $\tilde{N}(\lambda)$ 为其**分子**, $\tilde{M}(\lambda)$ 为其**分母**.

注意, $M(\lambda)$, $N(\lambda)$, $\tilde{M}(\lambda)$, $\tilde{N}(\lambda)$ 都是多项式矩阵,左 m.f.d.和右 m.f.d.是将通常有理函数表示成两个多项式之比推广于有理矩阵.这是换一种方式考察有理矩阵,在控制论中是有其用处的.

例 对于(7.3)中的矩阵 $A(\lambda)$ 有

$$\begin{aligned} g_1(\lambda)&=(\lambda+3)^2(\lambda+4), \quad & g_2(\lambda)&=(\lambda+3)(\lambda+4)^2; \\ \tilde{g}_1(\lambda)&=(\lambda+3)^2(\lambda+4)^2, \quad & \tilde{g}_2(\lambda)&=(\lambda+3)(\lambda+4)^2. \end{aligned}$$

$A(\lambda)$ 的左 m.f.d.为

$$\begin{aligned} A(\lambda)&=(M(\lambda))^{-1}N(\lambda) \\ &\equiv\begin{bmatrix} (\lambda+3)^2(\lambda+4) & 0 \\ 0 & (\lambda+3)(\lambda+4)^2 \end{bmatrix}^{-1} \end{aligned}$$

$$\times\begin{bmatrix}\lambda+4 & \lambda+3\\ -6 & (\lambda-1)(\lambda+3)\end{bmatrix}.$$

$A(\lambda)$的右 m.f.d.为

$$\begin{aligned}A(\lambda)&=\tilde{N}(\lambda)\left(\tilde{M}(\lambda)\right)^{-1}\\&\equiv\begin{bmatrix}(\lambda+4)^2 & \lambda+4\\ -6(\lambda+3) & (\lambda-1)(\lambda+3)\end{bmatrix}\\&\quad\times\begin{bmatrix}(\lambda+3)^2(\lambda+4)^2 & 0\\ 0 & (\lambda+3)(\lambda+4)^2\end{bmatrix}^{-1}.\end{aligned}$$

10 其它特殊矩阵综述

10.1 矩阵的有向图及指标矩阵

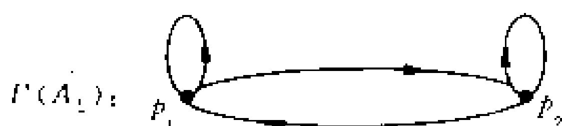
10.1.1 定义 设 $A=[a_{ij}]\in\mathbb{C}^{n\times n}$, $p_1,\cdots,p_n$ 是平面上任意 n 个不同的点,称其为**结点**(node).

对应 A 的每个非零元素 a_{ij},用一条以 p_i 为起点以 p_j 为终点的弧连接结点 p_i 和 p_j,并称其为**有向弧**(directed arc).

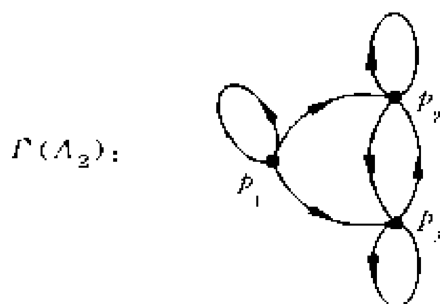
特别,对应 A 的非零元素 a_{ii},有向弧从 p_i 到其自身,称其为**闭路**(loop).

如此,矩阵 A 有具 n 个结点及有向弧组成的图,称为 A 的**有向图**(directed graph).记作 $\Gamma(A)$.

例

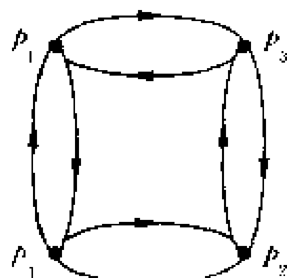


$$A_1=\begin{bmatrix}1&1\\1&1\end{bmatrix},$$



$$A_2=\begin{bmatrix}4&2&1\\0&1&1\\0&1&1\end{bmatrix}.$$

$$A_3 = \frac{1}{4}\begin{bmatrix} 0 & 0 & 1 & 1 \\ 0 & 0 & 1 & 1 \\ 1 & 1 & 0 & 0 \\ 1 & 1 & 0 & 0 \end{bmatrix}, \quad \Gamma(A_3):$$



10.1.2 定义 有向图Γ中的一条**有向路径**γ是指Γ中的一个有向弧序列

$$p_{i_1}p_{i_2}, p_{i_2}p_{i_3}, p_{i_3}p_{i_4}, \cdots \tag{1.1}$$

相应的有向路径γ的**结点有序表**是指

$$p_{i_1}, p_{i_2}, p_{i_3}, p_{i_4}, \cdots \tag{1.2}$$

有向路径γ的**长**是指γ中相继有向弧的数目,如果这个数目是有限的话;否则称γ是**无限长**的.

循环(cycle)或称**简单有向循环**(simple directed cycle)或**闭路径**(closed path)是指起点和终点在同一结点的有向路径;该结点正好在路径结点有序表中出现两次;在这个结点表中没有其它结点再次出现.长为1的循环即是闭路,也称为**平凡循环**(trivial cycle).

10.1.3 定义 有向图Γ称为**强连通**的,如果Γ中每对不同的结点p_i和p_j,存在以p_i为起点以p_j为终点的有限长有向路径.

10.1.4 定义 设$A=[a_{ij}]\in\mathbb{C}^{m\times n}$,$M(A)=[\mu_{ij}]\in\mathbb{C}^{m\times n}$称为$A$的**指标矩阵**(indicator matrix)或**关联矩阵**(incidence matrix),如果

$$\mu_{ij} = \begin{cases} 1, & a_{ij} \neq 0, \\ 0, & a_{ij} = 0, \end{cases} \qquad i, j = 1, \cdots, n. \tag{1.3}$$

显然,对于任何$A\in\mathbb{C}^{n\times n}$,$A$的有向图和其指标矩阵$M(A)$的

有向图是相同的.所以,考察 A 的有向图等价于考察 $M(A)$ 的有向图;有关的结论也因此常用两种等价的说法.

10.1.5 定理 设 $A\in\mathbb{C}^{n\times n}$, $A\geq 0$,且 A 是无穷可分的,则指标矩阵 $M(A)$ 是半正定的.

证 根据无穷可分的定义 **9.4.8**, $A\equiv[a_{ij}]$ 的 Hadamard 幂 $A^{(\alpha)}=\left[a_{ij}^{\alpha}\right]$ 对所有 $\alpha>0$ 是半正定的;再注意到

$$a_{ij}>0,\quad i,j=1,\cdots,n.$$

而且特征值连续地依赖于矩阵的元素,于是

$$M(A)=\lim_{\alpha\to 0}A^{(\alpha)}$$

必为半正定矩阵. □

10.1.6 定理 设 $A=[a_{ij}]\in\mathbb{C}^{n\times n}$, p_i 和 p_j 是 $\Gamma(A)$ 的结点,则 $\Gamma(A)$ 中 p_i 和 p_j 之间存在长为 m 的有向路径的充分必要条件是

$$\left(|A|^{m}\right)_{ij}\neq 0.$$

等价地说, $\Gamma(A)$ 中 p_i 和 p_j 之间存在长为 m 的有向路径的充分必要条件是

$$\left(M(A)^{m}\right)_{ij}\neq 0.$$

这里 $M(A)$ 是 A 的指标矩阵.

证 利用数学归纳法.

对于 $m=1$,结论显然成立.

为了获得推证线索,再看 $m=2$,由于

$$\left(|A|^{2}\right)_{ij}=\sum_{k=1}^{n}\left(|A|\right)_{ik}\left(|A|\right)_{kj}=\sum_{k=1}^{n}|a_{ik}||a_{kj}|,$$

可见, $\left(|A|^{2}\right)_{ij}\neq 0$ 当且仅当至少有一个 k, a_{ik} 和 a_{kj} 两者同不为零.然而如此情形当且仅当在 $\Gamma(A)$ 中存在从 p_i 到 p_j 的长为 2

的路径.

一般,假设对于 $m=q$ 结论成立，则

$$\left(|A|^{q+1}\right)_{ij}=\sum_{k=1}^{n}\left(|A|^{q}\right)_{ik}\left(|A|\right)_{kj}=\sum_{k=1}^{n}\left(|A|^{q}\right)_{ik}|a_{kj}|\neq 0,$$

当且仅当至少有一个 k ,$\left(|A|^{q}\right)_{ik}$ 和 a_{kj} 两者同时不为零.这等价于有一条从 p_i 到 p_k 的长为 q 的路径和有一条从 p_k 到 p_j 的长为1的路径.然而成立如此情形当且仅当存在一条从 p_i 到 p_j 的长为 $q+1$ 的路径.

关于 $M(A)$,论证步骤是相同的. □

10.1.7 定理 设 $A\in\mathbb{C}^{n\times n}$, A 的有向图 $\Gamma(A)$ 中任何两个结点存在长为 m 的有向路径的充分必要条件是 $|A|^{m}>0$.

等价地说,$\Gamma(A)$ 中任何两个结点存在长为 m 的有向路径的充分必要条件是 $M(A)^{m}\neq 0$.这里 $M(A)$ 是 A 的指标矩阵.

证 这是 **10.1.6** 的直接推论. □

10.1.8 定义 设 $A=[a_{ij}]\in\mathbb{C}^{n\times n}$ 分块为

$$A=\begin{bmatrix} A_{11} & A_{12} & \cdots & A_{1m} \\ A_{21} & A_{22} & \cdots & A_{2m} \\ \cdots & \cdots & \cdots & \cdots \\ A_{m1} & A_{m2} & \cdots & A_{mm} \end{bmatrix}, \tag{1.4}$$

其中

$$A_{ii}\in\mathbb{C}^{n_i\times n_i},\quad i=1,\cdots,m;\quad n_1+\cdots+n_m=n.$$

设 $p_{\pi(1)},\cdots,p_{\pi(m)}$ 是平面上的 m 个结点.对于 A 的分块形式(1.4)的每个非零子矩阵 A_{ij} ,用一条以 $p_{\pi(i)}$ 为起点以 $p_{\pi(j)}$ 为终点的

有向弧连接结点 $p_{\pi(i)}$ 和 $p_{\pi(j)}$.

如此,m 个结点及有向弧组成的图,称为 A 的相对于分块形式(1.4)的**块有向图**(block directed graph).记作 $\Gamma_\pi(A)$.

这是将矩阵的有向图概念 **10.1.1** 推广至分块情形.有向路径及其长,循环以及强连通等术语均可相应地推广并袭用.

下面引进无向图和邻接矩阵的概念.

10.1.9 定义 设

$$N \equiv \{p_1, \cdots, p_n\}$$

是元素称为**结点**的集合,而

$$E \equiv \{\{p_i, p_j\}: 1 \le i, j \le n\}$$

是由称为**边**的无序**结点对**组成的集合,N 和 E 两者在一起,称为一个**无向图**,记作 Γ.矩阵 $A = [a_{ij}] \in \mathbb{R}^{n\times n}$,

$$a_{ij} = \begin{cases} 1, & \{p_i, p_j\} \in E, \\ 0, & \{p_i, p_j\} \notin E, \end{cases} \quad i, j = 1, \cdots, n,$$

称为无向图 Γ 的**邻接矩阵**(adjacency matrix).

因为 $\{p_i, p_j\} = \{p_j, p_i\}$,$\Gamma$ 是无向的,所以邻接矩阵是实对称的非负矩阵.

10.2 性质 P 和性质 SC

10.2.1 定义 设 $A, B \in \mathbb{C}^{n\times n}$,而且

$$\lambda(A) = \{\lambda_1, \cdots, \lambda_n\}, \quad \lambda(B) = \{\mu_1, \cdots, \mu_n\},$$

这里特征值均按重数计.

如果对于任何$\alpha, \beta \in \mathbb{C}$,成立

$$\lambda(\alpha A+\beta B)=\{\alpha\lambda_j+\beta\mu_{i_j}: j=1,\cdots,n\}, \tag{2.1}$$

其中$i_1,\cdots,i_n$是指标$1,\cdots,n$的某一置换,则称A和B这一对矩阵具有**性质 L**.

如果对于任何二(非交换)元复系数多项式$p(t,s)$,成立

$$\lambda(p(A,B))=\{p(\lambda_j,\mu_{i_j}): j=1,\cdots,n\}, \tag{2.2}$$

其中$i_1,\cdots,i_n$是指标$1,\cdots,n$的某一置换,则称A和B这一对矩阵具有**性质 P**.

显然,性质 P 蕴涵性质 L,但反之不然.

例 一对正规矩阵$A, B \in \mathbb{C}^{n\times n}$具有性质 L,依 **5.1.9** 及 **2.4.4**, A和B必须可交换,因而必须是同时酉可对角化的.

对于性质 L 无足够了解.对于性质 P 有如下定理.

10.2.2 定理 设$A, B \in \mathbb{C}^{n\times n}$.存在非奇异矩阵$S \in \mathbb{C}^{n\times n}$使得

$$S^{-1}AS, \quad S^{-1}BS$$

同时为上三角矩阵的充分必要条件是A和B具有性质 P.

定理证明从略.

性质 L 和性质 P 的概念,上述定理及其推广的证明,可参见[MT1936]和[MT1952].

10.2.3 定义 $A \in \mathbb{C}^{n\times n}$称为具有**性质 SC**,如果对于每一有序对

$$p, q \in \{1,\cdots,n\}, \quad p \neq q,$$

存在由不同整数组成的序列

$$p=k_1,k_2,\cdots,k_{m-1},k_m=q \in \{1,\cdots,n\}, \tag{2.3}$$

使得 A 的元素

$$a_{k_1k_2}, a_{k_2k_3}, \cdots, a_{k_{m-1}k_m} \tag{2.4}$$

全不为零.

例 矩阵

$$A = \begin{bmatrix} 4 & 2 & 1 \\ 0 & 1 & 1 \\ 0 & 1 & 1 \end{bmatrix}$$

不具性质 SC.因为对于有序对2,1,不存在上述定义中那样的序列.然而对于有序对1,2,由于 $a_{12} = 2 \neq 0$,故1,2 本身就是上述定义中那样的序列.

10.2.4 定理 设 $A \in \mathbb{C}^{n\times n}$, A 具有性质 SC 的充分必要条件是 A 的有向图 $\Gamma(A)$ 是强连通的.

证明留作练习.

10.2.5 Better 定理 设 $A = [a_{ij}] \in \mathbb{C}^{n\times n}$, λ 是 A 的特征值而且是 Gerschgorin 域 $G(A)$ 的边界点.设 A 具有性质 SC,则

(1) A 的每个 Gerschgorin 圆通过 λ.

(2) 如果 $Ax = \lambda x$, $x = (x_1, \cdots, x_n)^{\mathrm{T}} \neq 0$,那么

$$|x_1| = \cdots = |x_n|.$$

证 设 $Ax = \lambda x$, $x = (x_1, \cdots, x_n)^{\mathrm{T}} \neq 0$ 且

$$|x_p| \equiv \|x\|_\infty > 0,$$

则依 **1.6.26** 的(1),成立

$$|\lambda - a_{pp}| = \sum_{j=1, j\neq p}^{n} |a_{pj}|.$$

表明 A 的第 p 个 Gerschgorin 圆通过 λ.

设 q 是另一个指标, $1 \le q \le n, q \neq p$. 因 A 具有性质 SC,故存在由不同指标组成的序列 $p = k_1, k_2, \cdots, k_{m-1}, k_m = q$, 使得 A 的元素 $a_{k_1k_2}, a_{k_2k_3}, \cdots, a_{k_{m-1}k_m}$ 全不为零.

依 **1.6.26** 的(2),由于 $a_{k_1k_2} = a_{pk_2} \neq 0$,有

$$\left|x_{k_2}\right| = \left|x_p\right|.$$

又由于 $a_{k_2k_3} \neq 0$,有

$$\left|x_{k_3}\right| = \left|x_{k_2}\right| = \left|x_p\right|.$$

如此继续,得

$$\left|x_{k_i}\right| = \left|x_p\right|, \quad i = 1, \cdots, m.$$

于是 $\left|x_q\right| = \left|x_p\right|$. 由此,仍依 **1.6.26** 的(1),成立

$$\left|\lambda - a_{k_mk_m}\right| = \left|\lambda - a_{qq}\right| = \sum_{j=1, j\neq q}^{n} \left|a_{qj}\right|.$$

也就是说, A 的第 q 个 Gerschgorin 圆通过 λ.

至此,从指标 q 的任意性即得结论(1)和(2). □

10.2.6 Better 推论 设 $A = [a_{ij}] \in \mathbb{C}^{n\times n}$ 具有性质 SC 而且是弱严格对角优势的,则 A 是可逆的.

证 利用反证法.

若 A 不可逆,则 0 是 A 的一个特征值. 因 A 是弱严格对角优势的, 0 不可能是 Gerschgorin 域 $G(A)$ 的内点,故必为边界点. 于是,依 **10.2.5**, A 的每个 Gerschgorin 圆通过 0;但是,至少有一个指标 i,成立

$$\left|a_{ii}\right| > \sum_{j=1, j\neq i}^{n} \left|a_{ij}\right|,$$

表明第 i 个 Gerschgorin 圆不可能通过 0. 矛盾. □

10.2.7 定理 设 $A\in\mathbb{C}^{n\times n}$, A 具有性质 SC 的充分必要条件是

$$\left(I+|A|\right)^{n-1}>0. \tag{2.5}$$

等价地说, A 具有性质 SC 的充分必要条件是

$$\left(I+M(A)\right)^{n-1}>0, \tag{2.6}$$

这里 $M(A)$ 是 A 的指标矩阵.

证 注意到有

$$\left(I+|A|\right)^{n-1}=I+(n-1)|A|+\binom{n-1}{2}|A|^2+\cdots+\binom{n-1}{n-1}|A|^{n-1}.$$

因此, $\left(I+|A|\right)^{n-1}>0$ 等价于对应每对指标

$$(i,j),\quad i\neq j,$$

在

$$|A|,|A|^2,\cdots,|A|^{n-1}$$

中,至少有一个其 (i,j)-元素是正的.

如此,依 **10.1.6**,又等价于 A 的有向图 $\Gamma(A)$ 中存在从 p_i 到 p_j 的有向路径.换句话说,等价于 $\Gamma(A)$ 是强连通的.

然而依 **10.2.4**, $\Gamma(A)$ 强连通等价于 A 具有性质 SC. □

10.2.8 推论 A 的有向图 $\Gamma(A)$ 中存在从 p_i 到 p_j 的有向路径的充分必要条件是

$$\left(\left(I+|A|\right)^{n-1}\right)_{ij}\neq 0.$$

证明留作练习.

10.2.9 定理 设 $A\in\mathbb{C}^{n\times n}$,下列条件等价:

(1) A 是不可约的.

(2) $\left(I+|A|\right)^{n-1}>0$.

(3) $(I+M(A))^{n-1}>0$,其中$M(A)$是A的指标矩阵.

(4) A的有向图$\Gamma(A)$是强连通的.

(5) A具有性质SC.

证 这是**4.2.1**,**10.2.4**和**10.2.7**的总和. □

10.2.10 定理 设$A\in\mathbb{C}^{n\times n}$是不可约的,而且至少有一个$i$,成立

$$R_i(A)\equiv\sum_{j=1}^{n}\left|a_{ij}\right|<\|A\|_\infty, \tag{2.7}$$

则

$$\rho(A)<\|A\|_\infty. \tag{2.8}$$

更一般地,若

$$D=\operatorname{diag}(d_1,\cdots,d_n),\quad d_1,\cdots,d_n>0,$$

而且至少有一个i,成立

$$R_i\left(D^{-1}AD\right)<\left\|D^{-1}AD\right\|_\infty,$$

则

$$\rho(A)<\left\|D^{-1}AD\right\|_\infty. \tag{2.9}$$

证 一般,

$$\rho(A)\le\|A\|_\infty,$$

其等号当且仅当A存在特征值λ满足

$$|\lambda|=\|A\|_\infty.$$

但此时,λ必是A的 Gerschgorin 域$G(A)$的边界点,依 **10.2.9** 和 **10.2.5**,A的每个Gerschgorin圆通过λ.然而,定理的假设(2.7)排除了这种可能.

将同样的论证应用于$D^{-1}AD$,即得

$$\rho(A)=\rho\left(D^{-1}AD\right)<\left\|D^{-1}AD\right\|_\infty. \qquad \square$$

10.3 性质 A 和 p-循环矩阵

10.3.1 定义 矩阵 $A=[a_{ij}]\in\mathbb{C}^{n\times n}$ 称为具有**性质** A(property A),如果存在子集

$$S,T\subset\{1,\cdots,n\};\ S\cap T=\emptyset,S\cup T=\{1,\cdots,n\},\quad(3.1)$$

使得对于任何 $a_{ij}\neq 0$ 或者有:

(1) $i=j$,或者

(2) $i\in S$ 且 $j\in T$,或者

(3) $i\in T$ 且 $j\in S$.

性质 A 的定义是 Young 于 1950 年旨在研究逐次超松弛(SOR)迭代法时提出的.取得的结果可以应用于出自一大类椭圆型偏微分方程在一般域上离散逼近所产生的矩阵方程的迭代求解.尔后,Arms,Gates 和 Zondek 于 1956 年给出性质 A^{π} 的定义,推广了 Young 的结果.

10.3.2 定义 矩阵 $A=[a_{ij}]\in\mathbb{C}^{n\times n}$ 称为具有**性质** A^{π} (property A^{π}),如果存在 A 的分块形式

$$A=\begin{bmatrix}A_{11}&A_{12}&\cdots&A_{1m}\\A_{21}&A_{22}&\cdots&A_{2m}\\\cdots&\cdots&\cdots&\cdots\\A_{m1}&A_{m2}&\cdots&A_{mm}\end{bmatrix},$$

其中

$$A_{ii}\in\mathbb{C}^{n_i\times n_i},\ i=1,\cdots,m;\quad n_1+\cdots+n_m=n,$$

使得 A 相对于这样的分块是具有性质 A 的.

10.3.3 定理 $A=[a_{ij}]\in\mathbb{C}^{n\times n}$ 具有性质 A 的充分必要条件是存在

$n\times n$ 置换矩阵 P,使得 A 的相合矩阵 $P^{\mathrm{T}}AP$ 形如

$$P^{\mathrm{T}}AP=\begin{bmatrix} D_1 & F \\ G & D_2 \end{bmatrix}, \tag{3.2}$$

其中 D_1 和 D_2 是对角矩阵.

定理证明从略.

关于性质 A,可参见[Y1954].

关于性质 A^{π},可参见[AGZ1956].

10.3.4 定义 设 $A=[a_{ij}]\in\mathbb{C}^{n\times n}$ 分块为

$$A=\begin{bmatrix} A_{11} & A_{12} & \cdots & A_{1m} \\ A_{21} & A_{22} & \cdots & A_{2m} \\ \cdots & \cdots & \cdots & \cdots \\ A_{m1} & A_{m2} & \cdots & A_{mm} \end{bmatrix}, \tag{3.3}$$

其中 $A_{ii}\in\mathbb{C}^{n_i\times n_i}$ 非奇异,$i=1,\cdots,m,n_1+\cdots+n_m=n$.记

$$D=\mathrm{diag}(A_{11},\cdots,A_{mm}).$$

相伴 A 的分块形式(3.3)的块 Jacobi 矩阵(见 **8.2.6**)是

$$J=I-D^{-1}A. \tag{3.4}$$

如果 J 是指数为 $p>1$ 的弱循环矩阵(见 **4.3.4**),则称 A 为相对于分块(3.3)的 **p-循环矩阵**(p-cyclic matrix).

例 两种特别关注的分块矩阵是形如

$$A_1=\begin{bmatrix} A_{11} & 0 & \cdots & 0 & A_{1p} \\ A_{21} & A_{22} & \ddots & & 0 \\ & \ddots & \ddots & \ddots & \vdots \\ & 0 & \ddots & \ddots & 0 \\ & & & A_{p,p-1} & A_{pp} \end{bmatrix}, p\geq 2 \tag{3.5}$$

的矩阵和**块三对角矩阵**

$$A_2=\begin{bmatrix} A_{11} & A_{12} & & \huge{0} \\ A_{21} & A_{22} & \ddots & \\ & \ddots & \ddots & A_{m-1,m} \\ \huge{0} & & A_{m,m-1} & A_{mm} \end{bmatrix}. \tag{3.6}$$

相伴的块 Jacobi 矩阵分别形如

$$J_1=\begin{bmatrix} 0 & \cdots & 0 & B_{1p} \\ B_{21} & 0 & & 0 \\ & \ddots & \ddots & \vdots \\ & & B_{p,p-1} & 0 \end{bmatrix} \tag{3.7}$$

和

$$J_2=\begin{bmatrix} 0 & B_{12} & & \\ B_{21} & 0 & \ddots & \\ & \ddots & \ddots & B_{m-1,m} \\ \huge{0} & & B_{m,m-1} & 0 \end{bmatrix}, \tag{3.8}$$

J_1 是指数为 p 的弱循环矩阵(见 **4.3.4**),故 A_1 是 p-循环矩阵.

可以证明(留作练习):J_2 是指数为 2 的弱循环矩阵,故 A_2 是 2-循环矩阵(提示:利用 J_2 的有向图).

10.3.5 定理　设 $A=[a_{ij}]\in\mathbb{C}^{n\times n}$ 的分块形式如同(3.3).

(1) 如果 $A_{ii}\equiv[a_{ii}]$ 即 1×1 矩阵,$i=1,\cdots,n$,那么 A 是 2-循环矩阵等价于 A 具有性质 A .

(2) 如果 A_{ii} 的阶没有进一步的限制,$i=1,\cdots,m$,那么 A 是 2-循环矩阵蕴涵 A 具有性质 A^{π} .

定理证明从略.

10.3.6 定理 设 $A=[a_{ij}]\in\mathbb{C}^{n\times n}$ 的分块形式是(3.3)中取 $m=2$ 的任一情况.如果 A 的对角块子矩阵是非奇异的,那么 A 是相对于如此分块的2-循环矩阵.

证 设

$$A=\begin{bmatrix}A_{11} & A_{12}\\ A_{21} & A_{22}\end{bmatrix},$$

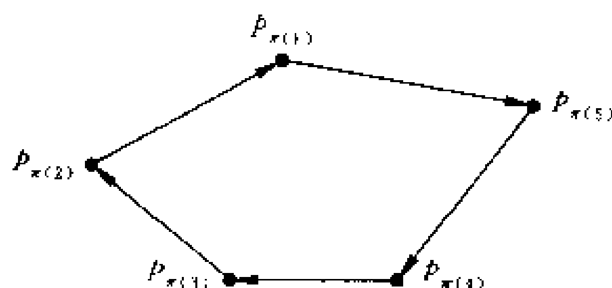
因 A_{11},A_{22} 非奇异,$D\equiv\operatorname{diag}(A_{11},A_{22})$可逆,故存在相伴的块 Jacobi 矩阵

$$J=I-D^{-1}A=\begin{bmatrix}0 & -A_{11}^{-1}A_{12}\\ -A_{22}^{-1}A_{21} & 0\end{bmatrix}.$$

依 **4.3.4**,J 显然是指数为 $p=2$ 的弱循环矩阵,因此 A 是2-循环矩阵. □

10.3.7 定义 设 $A\in\mathbb{C}^{n\times n}$ 的分块形式如同(3.3),而且其块有向图 $\Gamma_{\pi}(A)$ 是强连通的,则 $\Gamma_{\pi}(A)$ 的每个结点均存在循环.如果各循环之长的最大公因子为 p,那么称 $\Gamma_{\pi}(A)$ 是 A 相应于分块形式(3.3)的**指数为 p 的循环图**(cyclic graph).

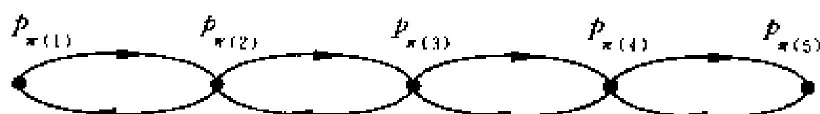
例 (3.7)的块矩阵 J_1,当 $p=5$ 时的块有向图 $\Gamma_{\pi}(J_1)$ 是



显然,$\Gamma_{\pi}(J_1)$ 是强连通的,且其每个结点的循环之长均为5,故

$\Gamma_\pi(J_1)$是指数为 $p=5$ 的循环图.

类似地,(3.8)的块矩阵 J_2,当 $m=5$ 时的块有向图 $\Gamma_\pi(J_2)$是



显然,$\Gamma_\pi(J_2)$也是强连通的,而且其每个结点的循环之长均为 2,故 $\Gamma_\pi(J_2)$是指数为 $p=2$ 的循环图.

10.3.8 定理 设(3.4)的块 Jacobi 矩阵 J 的块有向图 $\Gamma_\pi(J)$是强连通的.那么,若 $\Gamma_\pi(J)$是指数为 p 的循环图,则 A 是相对于分块(3.3)的 p-循环矩阵.

证明留作练习.

10.3.9 定义 设(3.3)的 A 是 p-循环矩阵.对于 $\alpha \neq 0$,引进从 Jacobi 矩阵 $J=L+U$ (见 **8.2.6**)导出的矩阵

$$J(\alpha) \equiv \alpha L + \alpha^{-(p-1)} U . \tag{3.9}$$

如果 $J(\alpha)$的所有特征值与 α 无关,则称 A 为**相容次序(p-循环)矩阵**(consistently ordered p–cyclic matrix).同时也说矩阵 J 是**相容次序**的.否则,称 A 和 J 都是**非相容次序**的.

例 考虑(3.5)的 A_1 和(3.6)的 A_2.(3.7)的 J_1 是 A_1 的块 Jacobi 矩阵,根据(3.9),注意到 A_1 是 p-循环矩阵,其导出矩阵为

$$J_1(\alpha) = \begin{bmatrix} 0 & \cdots & 0 & \alpha^{-(p-1)} B_{1p} \\ \alpha B_{21} & \ddots & & 0 \\ & \ddots & 0 & \vdots \\ \text{\Large 0} & & \alpha B_{p,p-1} & 0 \end{bmatrix},$$

直接计算得

$$J_1^p(\alpha)=J_1^p,\quad \forall\alpha\neq 0.$$

表明$J_1(\alpha)$的特征值与α无关,因此A_1是相容次序p-循环矩阵.

(3.8)的J_2是A_2的块 Jacobi 矩阵,根据(3.9),注意到A_2是2-循环矩阵,

$$p=2,\quad \alpha^{-(p-1)}=\alpha^{-1},$$

J_2的导出矩阵

$$J_2(\alpha)=\begin{bmatrix} 0 & \alpha^{-1}B_{12} & & \\ \alpha B_{21} & \ddots & \ddots & \\ & \ddots & 0 & \alpha^{-1}B_{m-1,m} \\ & & \alpha B_{m,m-1} & 0 \end{bmatrix}.$$

如果

$$J_2(\alpha)x=\lambda x,\tag{3.10}$$

其中$x=\left(x^{(1)},\cdots,x^{(m)}\right)^{\mathrm{T}}\neq 0$是与块矩阵相匹配的块向量,那么(3.10)等价于方程组

$$\alpha B_{i,i-1}x^{(i-1)}+\frac{1}{\alpha}B_{i,i+1}x^{(i+1)}=\lambda x^{(i)},\ i=1,\cdots,m,\tag{3.11}$$

其中B_{10}和$B_{m,m+1}$是零矩阵.现在令

$$z^{(i)}\equiv\frac{1}{\alpha^{i-1}}x^{(i)},$$

则(3.11)成为如下形式:

$$B_{i,i-1}z^{(i-1)}+B_{i,i+1}z^{(i+1)}=\lambda z^{(i)},\ i=1,\cdots,m.$$

这表明对于任何$\alpha\neq 0$,$J_2(\alpha)$的任一特征值是J_2的特征值,必与α无关,因此A_2是相容次序2-循环矩阵.换句话说,只要对角块子矩阵非奇异,任何形如(3.6)的块三对角矩阵必是相容次序2-循环矩阵.

关于p-循环矩阵和相容次序等概念应用于 SOR 迭代法的研

究,可参看[V1962].

10.4 素矩阵的有向图

10.4.1 定理 设 $A\in\mathbb{R}^{n\times n}$ 非负且不可约,$\{p_1,\cdots,p_n\}$ 是 A 的有向图 $\Gamma(A)$ 的结点集.用

$$L_i=\{k_1^{(i)},k_2^{(i)},\cdots\}$$

表示 $\Gamma(A)$ 中起点和终点都是结点 p_i 的所有有向路径长度之集,并用 g_i 表示 L_i 中所有长度的最大公因子,$i=1,\cdots,n$,则 A 为素矩阵的充分必要条件是

$$g_i=1,\quad i=1,\cdots,n. \tag{4.1}$$

证 由 A 不可约,易知 $L_i\neq\emptyset,i=1,\cdots,n$;而且依 **10.2.9**,对于每个 i 和任何 $j\neq i$,$\Gamma(A)$ 中存在从 p_i 到 p_j 的路径,也存在从 p_j 到 p_i 的路径.

如果 A 是素矩阵,则依 **4.3.9**,存在某 $m\geq 1$ 使得 $A^m>0$.由此并注意到 $A\geq 0$,得

$$A^k>0,\quad \forall k\geq m.$$

从而,依 **10.1.7**,有

$$m,m+1,m+2,\cdots\in L_i,\quad i=1,\cdots,n.$$

因此

$$g_i=1,\quad i=1,\cdots,n.$$

现在假设 A 不是素矩阵.如果 A 正好有 $k>1$ 个模最大的特征值,即 A 是指数为 k 循环矩阵,那么依 **4.3.3**,凡 m 不是 k 的整数倍,必有

$$a_{ii}^{(m)}=0,\quad i=1,\cdots,n,$$

这里 $A^{(m)} \equiv \left[a_{ii}^{(m)}\right]$.因此,依 **10.1.6**,即知

$$L_i \subset \{k, 2k, 3k, \cdots\}, \quad i = 1, \cdots, n,$$

从而

$$g_i \geq k > 1, \quad i = 1, \cdots, n. \qquad \square$$

10.4.2 定理 设 $A \in \mathbb{R}^{n \times n}$ 是素矩阵,则对某正整数

$$k \leq (n-1)n^n, \tag{4.2}$$

成立 $A^k > 0$.

证 因为 A 不可约,在 A 的有向图 $\Gamma(A)$ 中存在从结点 p_1 回到 p_1 的有向路径,最短的如此路径其长 $k_1 \leq n$.因此,依 **10.1.6**,矩阵 A^{k_1} 的 $(1,1)$-元素是正的,从而 A^{k_1} 的任何次幂的 $(1,1)$-元素也是正的.

注意到依 **4.3.7**, A^{k_1} 仍是素矩阵,必不可约,在 A^{k_1} 的有向图 $\Gamma\left(A^{k_1}\right)$ 中存在从结点 p_2 回到 p_2 的有向路径,最短的如此路径其长 $k_2 \leq n$.因此,矩阵

$$\left(A^{k_1}\right)^{k_2} = A^{k_1 k_2}$$

的 $(1,1)$-元素和 $(2,2)$-元素是正的.

依此类推,遍及 n 个主对角元素,得矩阵

$$A^{k_1 k_2 \cdots k_n}; \qquad k_i \leq n, \ i = 1, \cdots, n,$$

这个矩阵是不可约的且具有正对角元素.依 **4,3,8**,

$$\left(A^{k_1 k_2 \cdots k_n}\right)^{n-1} > 0,$$

其中

$$k \equiv k_1 k_2 \cdots k_n (n-1) \leq n^n (n-1). \qquad \square$$

10.4.3 定义 设 $A \in \mathbb{R}^{n \times n}$ 是素矩阵,使得 $A^k > 0$ 的最小的(正整数) k 称为 A 的**本原指标**(index of primitivity),记作 $\gamma(A)$.

4.3.8 表明当 A 的对角元素为正数时,

$$\gamma(A) \le n-1.$$

而一般情形,依 **10.4.2**,

$$\gamma(A) \le n^n(n-1),$$

但此界可以大大改进.

10.4.4 定理 设 $A \in \mathbb{R}^{n\times n}$ 是素矩阵,而且 A 的有向图 $\Gamma(A)$ 中最短简单有向循环其长为 s,则

$$A^{n+s(n-2)} > 0,$$

即有

$$\gamma(A) \le n+s(n-2). \tag{4.3}$$

定理证明从略.可参见[HJ1985]的 8.5.8.

10.4.5 推论 设 $A \in \mathbb{R}^{n\times n}$ 是非负矩阵,则 A 为素矩阵的充分必要条件是

$$A^{n^2-2n+2} > 0. \tag{4.4}$$

证 依 **4.3.9**,只要 A 的某个正整数次幂是正矩阵,A 必为素矩阵.故仅需证必要性.

设 A 是素矩阵,当 $n=1$ 时结论是显然的.

现设 $n>1$.因为 A 不可约,A 的有向图 $\Gamma(A)$ 中必存在循环.如果 $\Gamma(A)$ 中最短循环其长等于 n,那么 $\Gamma(A)$ 中每个循环之长是 n 的倍数,从而依 **10.4.1**,A 不是素矩阵.因此,$\Gamma(A)$ 中最短循环其长必 $\le n-1$,再依 **10.4.4**,得

$$\gamma(A) \le n+s(n-2) \le n+(n-1)(n-2) = n^2-2n+2.$$

由此推出(4.4)成立. □

10.4.6 定理 设 $A \in \mathbb{R}^{n\times n}$ 是不可约的非负矩阵,而且有 d 个正的主对角元素,$1 \le d \le n$.则

$$A^{2n-d-1} > 0,$$

即有

$$\gamma(A) \le 2n-d-1. \tag{4.5}$$

定理证明从略. 可参见[HJ1985]的 8.5.10.

10.4.7 定义 矩阵

$$A=\begin{bmatrix} 0 & 1 & 0 & \cdots & 0 \\ 0 & \ddots & \ddots & \ddots & \vdots \\ \vdots & \ddots & \ddots & \ddots & 0 \\ 0 & \cdots & 0 & 0 & 1 \\ 1 & 1 & 0 & \cdots & 0 \end{bmatrix} \in \mathbb{R}^{n\times n}$$

称为 **Wielandt 矩阵**.

10.4.8 定理 设 $A\in\mathbb{R}^{n\times n}$ 是 Wielandt 矩阵,则

(1) 当 $n\ge 3$ 时, A 是素矩阵.

(2) A^{n^2-2n+1} 的 $(1,1)$-元素等于零.

(3) $A^{n^2-2n+2}>0$.

证明留作练习.提示:(1)的证明可利用 A 的有向图 $\Gamma(A)$.(2)的证明可借助计算

$$Ae_i,\ \ A^{n-1}e_i,\ \ A^{(n-1)(n-1)}e_i,$$

e_i 是第 i 个分量为 1 的单位向量.

定理表明 Wielandt 矩阵给 **10.4.5** 提供了简明的实例.

10.4.9 定义 设 $A=[a_{ij}]\in\mathbb{R}^{n\times n}$ 是非负矩阵. A 称为**组合对称矩阵** (combinatorially symmetric matrix),如果

$$a_{ij}>0 \Leftrightarrow a_{ji}>0,\quad \forall i,j=1,\cdots,n.$$

10.4.10 定理 设 $A\in\mathbb{R}^{n\times n}$ 是组合对称的素矩阵,则 $A^{2n-2}>0$.

证明留作练习.提示:考虑 A^2,并利用 **4.3.7** 和 **10.4.6.**

10.5 初等矩阵

10.5.1 定义 设 $A\in\mathbb{C}^{m\times n}$.下列操作

(1) 在 A 中交换两行(列);

(2) A 的一行(列)的每个元素同乘以某非零数;

(3) A 的任一行(列)乘以某非零数加到 A 的另一行(列),

分别称为关于矩阵 A 的行(列)的 1,2,3 型**初等运算**.

关于矩阵 A 的行(列)初等运算可以通过对 A 乘以以下三类矩阵来实现:

对 A 左乘 $m\times m$(右乘 $n\times n$)矩阵

$$E^{(1)}=\begin{bmatrix}1&&&&&&&&&\\&\ddots&&&&&&&&\\&&1&&&&&&&\\&&&0&\cdots&\cdots&0&1&&\\&&&\vdots&1&&&0&&\\&&&\vdots&&\ddots&&\vdots&&\\&&&0&&&1&\vdots&&\\&&&1&0&\cdots&\cdots&0&&\\&&&&&&&&1&&\\&&&&&&&&&\ddots&\\&&&&&&&&&&1\end{bmatrix}\begin{matrix}\\\\\\i\\\\\\\\j\\\\\\\\\end{matrix}$$

实现交换 A 的第 i 行(列)和第 j 行(列).

对 A 左乘 $m\times m$(右乘 $n\times n$)矩阵

$$E^{(2)}=\begin{bmatrix}1 & & & & & & \\ & \ddots & & & & \mbox{\Large 0} & \\ & & 1 & & & & \\ & & & c & & & \\ & & & & 1 & & \\ & \mbox{\Large 0} & & & & \ddots & \\ & & & & & & 1\end{bmatrix} i$$

实现 A 的第 i 行(列)乘以数 c.

对 A 左乘 $m\times m$(右乘 $n\times n$)矩阵

$$E^{(3)}=\begin{bmatrix}1 & & & & \\ & \ddots & \cdots & c & \\ & & \ddots & \vdots & \\ \mbox{\Large 0} & & & \ddots & \\ & & & & 1\end{bmatrix}\begin{matrix} \\ i \\ \\ j \\ \\ \end{matrix} \quad 或 \quad E^{(3)}=\begin{bmatrix}1 & & & & \\ & \ddots & & \mbox{\Large 0} & \\ & \vdots & \ddots & & \\ & c & \cdots & \ddots & \\ & & & & 1\end{bmatrix}\begin{matrix} \\ j \\ \\ i \\ \\ \end{matrix}$$

实现 A 的第 j 行(列)乘以数 c 加到 A 的第 i 行(列).

$E^{(1)},E^{(2)},E^{(3)}$ 分别称为 1,2,3 型**初等矩阵**.

9.6.8 定义的初等变换矩阵蕴涵了初等矩阵.

10.5.2 定理 任何 $A\in\mathbb{C}^{m\times n}$ 通过行初等运算可以约化为具有如下性质的矩阵:

(1) 如果有元素全为零的行,它们都在矩阵的底部位置.

(2) 非零行中从左数起的第一个非零元素等于1(称为**首一**).

(3) 对于 $i=2,\cdots,m$,第 i 行的首一(如果有的话)出现在第 $i-1$ 行的首一之右.

(4) 任何包含首一的列其所有其它元素全为零.

证明留作练习.

具有定理中四条性质的矩阵称为**约化行梯矩阵**(reduced row echelon matrix).

例 下列矩阵都是约化行梯矩阵:

$$\begin{bmatrix}1 & 0 & -3\\0 & 1 & 2\end{bmatrix},\ \begin{bmatrix}1 & 0\\0 & 1\\0 & 0\end{bmatrix},\ \begin{bmatrix}1 & -2 & 0\\0 & 0 & 1\\0 & 0 & 0\end{bmatrix},\ \begin{bmatrix}1 & 2 & 1\\0 & 0 & 0\\0 & 0 & 0\end{bmatrix}.$$

显然,这些矩阵不可能通过行初等运算约化为更简单的形式.

但是,如果还允许列初等运算,那么正如下一定理所述那样,有可能进一步约化.

10.5.3 定理 任何 $A \in \mathbb{C}^{m\times n}$ 通过行和列初等运算可以约化为如下诸 $m\times n$ 矩阵之一:

$$[I_m \quad 0],\ \begin{bmatrix}I_n\\0\end{bmatrix},\ \begin{bmatrix}I_r & 0\\0 & 0\end{bmatrix},\ I_n(m=n), \tag{5.1}$$

其中 $r<\min\{m,n\}$.

证明留作练习.

注意到初等运算可以应用左乘或右乘初等矩阵来实现.上述定理有如下等价的陈述.

10.5.4 定理 设 $A \in \mathbb{C}^{m\times n}$,则存在初等矩阵有限序列

$$E_1,\cdots,E_{k+s},$$

使得

$$E_k\cdots E_1AE_{k+1}\cdots E_{k+s} \tag{5.2}$$

是(5.1)中诸矩阵之一.

10.5.5 推论 设 $A\in\mathbb{C}^{m\times n}$,则存在非奇异矩阵 $P\in\mathbb{C}^{m\times m}$ 和 $Q\in\mathbb{C}^{n\times n}$,使得 PAQ 是(5.1)中诸矩阵之一.

证 在(5.2)中取

$$P\equiv E_k\cdots E_1,\quad Q\equiv E_{k+1}\cdots E_{k+s}.$$ □

10.5.6 定理 设 $A,B\in\mathbb{C}^{m\times n}$,下列条件等价:

(1) A 和 B 有相同的(5.1)中的约化形式.

(2) 应用行和列初等运算,可从 A 和 B 的任一个得出另一个.

(3) 存在非奇异矩阵 $P\in\mathbb{C}^{m\times m}$ 和 $Q\in\mathbb{C}^{n\times n}$ 使得 $A=PBQ$.

证 存在初等矩阵 $E_1,\cdots,E_{k+s}$ 和 $\tilde{E}_1\cdots\tilde{E}_{l+t}$,使得

$$E_k\cdots E_1AE_{k+1}\cdots E_{k+s}=\tilde{E}_l\cdots\tilde{E}_1B\tilde{E}_{l+1}\cdots\tilde{E}_{l+t}$$

等价于

$$A=\left(E_1^{-1}\cdots E_k^{-1}\tilde{E}_l\cdots\tilde{E}_1\right)B\left(\tilde{E}_{l+1}\cdots\tilde{E}_{l+t}E_{k+s}^{-1}\cdots E_{k+1}^{-1}\right)\equiv PBQ,$$

或写成

$$B=\left(\tilde{E}_1^{-1}\cdots\tilde{E}_l^{-1}E_k\cdots E_1\right)A\left(E_{k+1}\cdots E_{k+s}\tilde{E}_{l+t}^{-1}\cdots\tilde{E}_{l+1}^{-1}\right)\equiv\tilde{P}A\tilde{Q},$$

而且,**10.5.4** 蕴涵了每个非奇异矩阵可以表示为有限个初等矩阵之积. □

10.5.7 定义 设 $A,B\in\mathbb{C}^{m\times n}$.如果存在非奇异矩阵 $P\in\mathbb{C}^{m\times m}$ 和 $Q\in\mathbb{C}^{n\times n}$,使得

$$A=PBQ,$$

则称矩阵 A 和 B **等价**,记作 $A\sim B$.

这是用 **10.5.6** 的条件(3)定义等价概念.从 **10.5.6** 可知,也可用其条件(1)或条件(3)作定义.

显然,矩阵等价概念是 $\mathbb{C}^{m\times n}$ 上的一个等价关系,即成立

(1) 自反性:$A\sim A$.

(2) 对称性:若 $A\sim B$,则 $B\sim A$.

(3) 传递性:若 $A \sim B$, $B \sim C$,则 $A \sim C$.

证明留作练习.

因此,对于固定的 m 和 n,可将$\mathbb{C}^{m\times n}$ 中的矩阵全体分裂成不相交的集合,每个集合是由彼此等价的矩阵组成的,称为**等价类**; $\mathbb{C}^{m\times n}$ 中的每个矩阵属于一个而且仅一个等价类.

每个等价类可以由形如(5.1)的矩阵表示和唯一确定,称为该类矩阵的**最简标准形**.

10.5.8 定理 设 $A, B \in \mathbb{C}^{m\times n}$, A 和 B 等价的充分必要条件是两者有相同的秩.

证明留作练习.

10.6 相合矩阵

10.6.1 定义 设 $A, B \in \mathbb{C}^{n\times n}$.如果存在非奇异矩阵 $S \in \mathbb{C}^{n\times n}$,使得

$$B = SAS^*, \tag{6.1}$$

则称 B 关于 A 是 ***相合**的(*congruent)或**共轭相合**的(conjunctive),也说 B 为 ***相合**或**共轭相合**于 A;或者,使得

$$B = SAS^{\mathrm{T}}, \tag{6.2}$$

则称 B 关于 A 是 ^{T}**相合**的(^{T}congruent),也说 B 为 ^{T}**相合**于 A.

从 A 到 SAS^* 或 SAS^{T} 统称为**相合变换**.

当 $S \in \mathbb{R}^{n\times n}$ 时,*相合和 ^{T}相合是一样的.

实或复二次型 $q(x) \equiv (x, Ax)$ 在坐标变换 $x = Sy$ 下,

$$q(Sy) = (Sy, ASy) = (y, By),$$

其中 $B \equiv S^* AS$ 是 ^{T} 相合或 * 相合于 A 的.

10.6.2 定理 * 相合和 ^{T} 相合都是等价关系.也就是说,对于任何 $A,B,C \in \mathbb{C}^{n\times n}$,有

(1) A 相合于 A.

(2) 若 A 相合于 B,则 B 相合于 A.

(3) 若 A 相合于 B, B 相合于 C,则 A 相合于 C.

证 对于 * 相合而言:

(1) $A = IAI^*$.

(2) 若 $A = SBS^*$, S 非奇异,则

$$B = S^{-1}A\left(S^{-1}\right)^*.$$

(3) 若 $A = S_1BS_1^*$, $B = S_2CS_2^*$,则

$$A = \left(S_1S_2\right)C\left(S_1S_2\right)^*.$$

对于 ^{T} 相合的证明,形式上是一样的. □

10.6.3 定理 设 $A,B \in \mathbb{C}^{n\times n}$ 是自伴矩阵,则 A 为 * 相合于 B 的充分必要条件是 A 和 B 有相同的惯性.

证 A 和 B 都是自伴矩阵.

具体地说,需要证明:存在非奇异矩阵 $S \in \mathbb{C}^{n\times n}$ 使得 $A = SBS^*$ 的充分必要条件是 A 和 B 两者正、负和零的特征值的个数分别相同.

先考虑自伴矩阵在相合变换下的表示.依 **3.2.3**,自伴矩阵酉等价于对角矩阵.

因此,对于 A,存在酉矩阵 $U \in \mathbb{C}^{n\times n}$,使得

$$A = U\Lambda U^*, \tag{6.3}$$

其中 $\Lambda = \mathrm{diag}\left(\lambda_1,\cdots,\lambda_n\right)$,而且 $\lambda_1,\cdots,\lambda_n \in \mathbb{R}$.为了方便,假设

$$\lambda_1,\cdots,\lambda_{i_1} > 0,8$$

$$\lambda_{i_+ +1},\cdots,\lambda_{i_+ + i_-} < 0, \tag{6.4}$$
$$\lambda_{i_+ + i_- +1} = \cdots = \lambda_n = 0.$$

令

$$D = \operatorname{diag}\left(\sqrt{\lambda_1},\cdots,\sqrt{\lambda_{i_+}},\sqrt{-\lambda_{i_+ +1}},\cdots\sqrt{-\lambda_{i_+ + i_-}},1,\cdots,1\right) \in \mathbb{R}^{n\times n}, \tag{6.5}$$

则

$$\Lambda = DI(A)D, \tag{6.6}$$

其中

$$I(A) \equiv \begin{bmatrix} 1 & & & & & & & & \\ & \ddots & & & & & & & \\ & & 1 & & & & 0 & & \\ & & & -1 & & & & & \\ & & & & \ddots & & & & \\ & & & & & -1 & & & \\ & & 0 & & & & 0 & & \\ & & & & & & & \ddots & \\ & & & & & & & & 0 \end{bmatrix}$$

称为 A 的**惯性矩阵**(inertia matrix),其对角元素正好有

i_+ 个“1”, i_- 个“-1”, i_0 个“0”.

$i(A) \equiv \{i_+, i_-, i_0\}$ 是 A 的惯性.于是,由(6.3)和(6.6), A 在相合变换下有如下表示:

$$A = SI(A)S^*, \tag{6.7}$$

其中 $S \equiv UD$,是非奇异矩阵.

现在设 A 和 B 有相同的惯性,则两者有形如(6.7)的表示,彼此的 S 不同,但有相同的惯性矩阵.因为 * 相合关系有传递性,而且 A 和 B 均 * 相合于同一矩阵,所以 A 和 B 是 * 相合的.

反之,设 A 和 B 是 * 相合的, $A = SBS^*$, S 非奇异.显然,相合矩阵有相同的秩,故

$$i_0(A)=i_0(B),$$

即 A 和 B 的零特征值的个数相同.从而只须再证明

$$i_+(A)=i_+(B),$$

即 A 和 B 的正特征值的个数也相同.

取 $v_1,\cdots,v_{i_+(A)}$ 是 A 的分别相应于正特征值 $\lambda_1,\cdots,\lambda_{i_+(A)}$ 的正交的特征向量,令

$$S_+(A)\equiv\operatorname{span}\left\{v_1,\cdots,v_{i_+(A)}\right\},$$

$\mathbb{C}^n$ 的线性子空间 $S_+(A)$ 的维数为 $i_+(A)$.如果

$$x=\alpha_1v_1+\cdots+\alpha_{i_+(A)}v_{i_+(A)}\neq 0,$$

那么

$$x^*Ax=\lambda_1|\alpha_1|^2+\cdots+\lambda_{i_+(A)}\left|\alpha_{i_+(A)}\right|^2>0.$$

注意到 $A=SBS^*$,有

$$x^*SBS^*x=(S^*x)^*B(S^*x)>0.$$

换言之,

$$y^*By>0,\quad \forall y\in\operatorname{span}\left\{S^*v_1,\cdots,S^*v_{i_+(A)}\right\},y\neq 0.$$

由此,并鉴于 $\operatorname{span}\left\{S^*v_1,\cdots,S^*v_{i_+(A)}\right\}$ 的维数也为 $i_+(A)$,推出

$$i_+(B)\geq i_+(A).$$

然而 A 和 B 的地位是对等的,故又成立

$$i_+(A)\geq i_+(B).$$ □

10.6.4 引理 设 $A\in\mathbb{C}^{n\times n}$ 是正定矩阵,$C\in\mathbb{C}^{n\times m}$,则 C^*AC 是半正定矩阵,而且

$$\operatorname{rank}(C^*AC)=\operatorname{rank}C.\tag{6.8}$$

这样,当 A 正定时,C^*AC 正定的充分必要条件是 $\operatorname{rank}C=m$.

证 首先注意 $C^*AC\in\mathbb{C}^{m\times m}$ 而且是自伴的.由 A 正定,有

$$x^*C^*ACx=y^*Ay\geq 0,\quad \forall x\in\mathbb{C}^m, y\equiv Cx,$$

故 C^*AC 是半正定的;而且,$x^*C^*ACx > 0$ 当且仅当 $Cx \neq 0$.

因此,若 $C^*ACx = 0$,则 $x^*C^*ACx = 0$,推出 $Cx = 0$.

反之,若 $Cx = 0$,则显然 $C^*ACx = 0$.

这样,$C^*ACx = 0$ 当且仅当 $Cx = 0$.这说明 C^*AC 和 C 不仅有相同的秩,而且有相同的零空间. □

10.6.5 定理 设 $A \in \mathbb{C}^{n\times n}$ 是半正定矩阵,$k \geq 1$ 是给定整数,则存在唯一的半正定矩阵 $B \in \mathbb{C}^{n\times n}$,使得

$$B^k = A.$$

并且满足

(1) $BA = AB$;

(2) $\mathrm{rank}B = \mathrm{rank}A$,故若 A 正定,则 B 也正定;

(3) 若 $A \in \mathbb{R}^{n\times n}$,则 $B \in \mathbb{R}^{n\times n}$.

证 由于 A 自伴,依 **3.2.3**,可以酉对角化,即存在酉矩阵 $U \in \mathbb{C}^{n\times n}$,使得

$$A = U\Lambda U^*, \quad \Lambda \equiv \mathrm{diag}(\lambda_1, \cdots, \lambda_n).$$

而且从 A 半正定知 $\lambda_1, \cdots, \lambda_n \geq 0$.定义

$$B = U\Lambda^{1/k}U^*, \quad \Lambda^{1/k} \equiv \mathrm{diag}\left(\lambda_1^{1/k}, \cdots, \lambda_n^{1/k}\right),$$

显然,$B^k = A$,B 是半正定的,$\mathrm{rank}B = \mathrm{rank}A$.而且

$$AB = U\Lambda U^*U\Lambda^{1/k}U^* = U\Lambda\Lambda^{1/k}U^* = U\Lambda^{1/k}\Lambda U^*$$

$$= U\Lambda^{1/k}U^*U\Lambda U^* = BA.$$

若 $A \in \mathbb{R}^{n\times n}$ 是半正定的,则 U 可取实正交矩阵.此时,显然可取 $B \in \mathbb{R}^{n\times n}$.

剩下的是唯一性问题,证明从略. □

10.6.6 定理 矩阵 $A \in \mathbb{C}^{n\times n}$ 正定的充分必要条件是 A 为 * 相合于单位矩阵 I.换言之,A 正定的充分必要条件是存在非奇异矩阵 $C \in \mathbb{C}^{n\times n}$ 使得

$$A = C^*C.$$

证 设 $A = C^*C$,矩阵 C 是非奇异的.

因 I 是正定的,依 **10.6.4**,$C^*C = C^*IC$ 是正定的.

反之,设 A 正定.则 $C \equiv A^{1/2}$ 存在且正定,$A = C^*C$. □

10.6.7 推论 矩阵 $A \in \mathbb{C}^{n\times n}$ 正定的充分必要条件是存在非奇异的具有正对角元素的上三角矩阵 $R \in \mathbb{C}^{n\times n}$,使得

$$A = R^*R, \tag{6.9}$$

当 $A \in \mathbb{R}^{n\times n}$ 时,可取 $R \in \mathbb{R}^{n\times n}$.

证 依 **10.6.6**,A 正定当且仅当 $A = C^*C$,C 非奇异.再仿 **5.7.5** 的推证,C 有 QR 分解,

$$C = QR,$$

其中 Q 是酉矩阵,R 是上三角矩阵,且可以选取所有对角元素为正的.于是

$$A = C^*C = (QR)^*(QR) = R^*Q^*QR = R^*R.$$

当 $A \in \mathbb{R}^{n\times n}$ 时,可取 $C \in \mathbb{R}^{n\times n}$,从而也可取 $Q, R \in \mathbb{R}^{n\times n}$. □

10.6.8 定理 设 $A \in \mathbb{C}^{n\times n}$ 是正定矩阵,$B \in \mathbb{C}^{n\times n}$ 是自伴矩阵,则 AB 是可对角化矩阵,其特征值全为实数. AB 和 B 有相同的惯性.

反之,任何特征值为实数的可对角化矩阵可以表示为正定矩阵和自伴矩阵之积.

证 A 正定,存在正定的平方根 $A^{1/2}$.注意到

$$A^{-1/2}ABA^{1/2} = A^{1/2}BA^{1/2},$$

这表明矩阵 AB 相似于 $A^{1/2}BA^{1/2}$,两者有相同的特征值.因 $A^{1/2}$ 自伴,矩阵 $A^{1/2}BA^{1/2}$ 相合于 B.依 **10.6.3**,B 和 $A^{1/2}BA^{1/2}$,从而和 AB 有相同的惯性.鉴于 $A^{1/2}BA^{1/2}$ 自伴,可对角化,故 AB 也必可对角化.

最后,如果 $C \in \mathbb{C}^{n\times n}$ 可对角化,而且特征值全为实数,即存在非奇异矩阵 $S \in \mathbb{C}^{n\times n}$ 和对角矩阵 $D \in \mathbb{R}^{n\times n}$,使得

$$C = SDS^{-1},$$

那么

$$C = S\left(S^*\left(S^*\right)^{-1}\right)DS^{-1} = \left(SS^*\right)\left(\left(S^{-1}\right)^* DS^{-1}\right) = AB,$$

其中

$$A \equiv SS^*$$

是正定的,

$$B \equiv \left(S^{-1}\right)^* DS^{-1}$$

是自伴的. □

10.6.9 定理 设 $A, B \in \mathbb{C}^{n\times n}$ 是两个自伴矩阵,而且存在 $\alpha, \beta \in \mathbb{R}$,使得 $P \equiv \alpha A + \beta B$ 是正定的,则存在非奇异矩阵 $C \in \mathbb{C}^{n\times n}$,使得 C^*AC 和 C^*BC 同为对角矩阵.

证 显然,α 和 β 至少有一个不为零.不妨设 $\beta \neq 0$,此时,

$$B = \beta^{-1}(P - \alpha A).$$

依 **10.6.6**, P 必 * 相合于单位矩阵,即存在非奇异矩阵 $\hat{C} \in \mathbb{C}^{n\times n}$ 使得 $\hat{C}^* P\hat{C} = I$.再注意到 $\hat{C}^* A\hat{C}$ 是自伴的,存在酉矩阵 $U \in \mathbb{C}^{n\times n}$ 使得

$$U^*\hat{C}^* A\hat{C}U \equiv D$$

是对角矩阵.于是,令

$$C \equiv \hat{C}U,$$

有

$$C^*PC = I, \qquad C^*AC = D.$$

并因此

$$C^*BC = \beta^{-1}(I - \alpha D)$$

也是对角矩阵. □

10.6.10 推论 设 $A \in \mathbb{C}^{n\times n}$ 是正定矩阵, $B \in \mathbb{C}^{n\times n}$ 是自伴矩阵,则存在非奇异矩阵 $C \in \mathbb{C}^{n\times n}$,使得 C^*BC 为对角矩阵,而且

$$C^*AC = I.$$ □

10.6.11 定理 设 $A \in \mathbb{C}^{n\times n}$ 是正定矩阵, $B \in \mathbb{C}^{n\times n}$ 是复对称矩阵, 则存在非奇异矩阵 $C \in \mathbb{C}^{n\times n}$, 使得 C^*AC 和 $C^{\mathrm{T}}BC$ 同为对角矩阵.

证 依 **10.6.6**, 可选取非奇异矩阵 $\hat{C} \in \mathbb{C}^{n\times n}$ 使得 $\hat{C}^*A\hat{C} = I$. 然后, 因 $\hat{C}^{\mathrm{T}}B\hat{C}$ 是对称的, 依 **10.7.3**, 存在酉矩阵 $U \in \mathbb{C}^{n\times n}$ 使得

$$U^{\mathrm{T}}\left(\hat{C}^{\mathrm{T}}B\hat{C}\right)U \equiv D$$

是对角矩阵. 于是, 再注意到

$$U^*\hat{C}^*A\hat{C}U = I,$$

可取 $C \equiv \hat{C}U$. □

10.6.12 定理 由

$$f(A) \equiv \log(\det A)$$

定义的 $f(\cdot)$ 是 $\mathbb{C}^{n\times n}$ 中正定矩阵凸集上的严格凹函数. 具体地说, 对于任何两个正定矩阵 $A, B \in \mathbb{C}^{n\times n}$, 成立

$$f(\alpha A + (1-\alpha)B) \geq \alpha f(A) + (1-\alpha)f(B),\ \forall \alpha \in (0,1), \tag{6.10}$$

当且仅当 $A = B$ 时取等号.

证 设 $A, B \in \mathbb{C}^{n\times n}$ 是正定矩阵. 依 **10.6.10**, 存在非奇异矩阵 $C \in \mathbb{C}^{n\times n}$, 使得

$$A = CIC^*, \qquad B = C\Lambda C^*,$$

其中

$$\Lambda \equiv \mathrm{diag}(\lambda_1, \cdots, \lambda_n), \quad \lambda_1, \cdots, \lambda_n > 0.$$

于是, 有

$$\begin{aligned} f(\alpha A + (1-\alpha)B) &= f\left(C\left[\alpha I + (1-\alpha)\Lambda\right]C^*\right) \\ &= f(CC^*) + f(\alpha I + (1-\alpha)\Lambda) \\ &= f(A) + f(\alpha I + (1-\alpha)\Lambda) \end{aligned} \tag{6.11}$$

和

$$
\begin{aligned}
\alpha f(A)+(1-\alpha)f(B) &= \alpha f(A)+(1-\alpha)f\left(C\Lambda C^*\right) \\
&= \alpha f(A)+(1-\alpha)\left[f\left(CC^*\right)+f(\Lambda)\right] \\
&= \alpha f(A)+(1-\alpha)f(A)+(1-\alpha)f(\Lambda) \\
&= f(A)+(1-\alpha)f(\Lambda). \qquad (6.12)
\end{aligned}
$$

再利用通常一元对数函数的严格凹性,有

$$
\begin{aligned}
f\left(\alpha I+(1-\alpha)\Lambda\right) &= \log\prod_{i=1}^{n}\left(\alpha+(1-\alpha)\lambda_i\right) \\
&= \sum_{i=1}^{n}\log\left(\alpha+(1-\alpha)\lambda_i\right) \\
&\geq \sum_{i=1}^{n}\left(\alpha\log 1+(1-\alpha)\log\lambda_i\right) \\
&= (1-\alpha)\sum_{i=1}^{n}\log\lambda_i = (1-\alpha)\log\prod_{i=1}^{n}\lambda_i \\
&= (1-\alpha)\log\left(\det\Lambda\right) = (1-\alpha)f(\Lambda), \\
&\qquad \forall\alpha\in(0,1).
\end{aligned}
$$

由此及(6.11)和(6.12)即得不等式(6.10).

(6.10)的等号成立当且仅当

$$\lambda_1 = \cdots = \lambda_n = 1,$$

亦即当且仅当 $\Lambda = I$,此时

$$B = CIC^* = A. \qquad \square$$

10.6.13 推论 设 $A,B\in\mathbb{C}^{n\times n}$ 是正定矩阵,则

$$\det\left[\alpha A+(1-\alpha)B\right]\geq\left(\det A\right)^{\alpha}\left(\det B\right)^{1-\alpha},\quad\forall\alpha\in(0,1),$$

当且仅当 $A = B$ 时取等号. $\square$

10.7 复对称矩阵

10.7.1 引言 矩阵 $A\in\mathbb{C}^{n\times n}$ 对称,即满足 $A=A^{\mathrm{T}}$.在许多场合中,所需研究的对称矩阵其元素全是实数,故属于实自伴矩阵,无须再专门讨论.相对来说,**复对称矩阵**——一般不是自伴矩阵——在应用中出现远不如实自伴矩阵那样经常,但是仍然会遇到,不乏应用的例子.

复对称矩阵与实对称矩阵的显著区别之一是不一定能对角化.例如,矩阵

$$A=\begin{bmatrix}1 & \mathrm{i}\\ \mathrm{i} & -1\end{bmatrix}$$

是不可对角化的.事实上,$A^2=0$,如果存在非奇异矩阵 $P\in\mathbb{C}^{2\times 2}$ 和对角矩阵 $\Lambda\in\mathbb{C}^{2\times 2}$,使得 $A=P\Lambda P^{-1}$,那么

$$0=A^2=\left(P\Lambda P^{-1}\right)\left(P\Lambda P^{-1}\right)=P\Lambda^2P^{-1},$$

推出 $\Lambda=0$,从而 $A=P\Lambda P^{-1}=0$.矛盾.

但是,对于复对称矩阵仍然有类似于自伴矩阵谱定理那样的分解.

10.7.2 定理 设 $A\in\mathbb{C}^{n\times n}$,存在酉矩阵 $U\in\mathbb{C}^{n\times n}$ 和上三角矩阵 $\Delta\in\mathbb{C}^{n\times n}$ 使得

$$A=U\Delta U^{\mathrm{T}}\tag{7.1}$$

的充分必要条件是 $A\overline{A}$ 的所有特征值为实的和非负的,在如此条件下,Δ 的所有主对角元素可以选取为非负的.

证 如果 $A=U\Delta U^{\mathrm{T}}$,U 是酉矩阵,Δ 是上三角矩阵,那么 $\overline{U}$ 是酉矩阵,$U^{\mathrm{T}}=\overline{U}^*$,得

$$A\overline{A}=U\Delta U^{\mathrm{T}}\overline{U\Delta U^*}=U\Delta\overline{\Delta}U^*$$

而且 $\Delta\overline{\Delta}$ 是上三角矩阵,其主对角元素是非负实数.于是,从 $A\overline{A}$ 相似

于 $A\overline{A}$ 以及上三角矩阵的特征值正好是它的主对角元素,必要性得证.

反之,假定 $A\overline{A}$ 的所有特征值是非负的.设 x 是 $A\overline{A}$ 的相应于特征值 $\lambda \geq 0$ 的特征向量,

$$A\overline{A}x = \lambda x.$$

存在两种可能: $A\overline{x}$ 与 x 线性相关;或者, $A\overline{x}$ 与 x 线性无关.

先看 $A\overline{x}$ 与 x 线性相关.此时,存在 $\mu \in \mathbb{C}$,使得 $A\overline{x} = \mu x$.从而

$$A\overline{A}x = \overline{\mu}A\overline{x} = \overline{\mu}\mu x = |\mu|^2 x = \lambda x,$$

得 $|\mu|^2 = \lambda$.注意,当 λ 是 $A\overline{A}$ 的单特征值时, $A\overline{x}$ 必与 x 线性相关.事实上,令 $\sigma = +\sqrt{\lambda}$, $w = A\overline{x} - \sigma x$,则因 $A\overline{A}x = \lambda x$,有

$$A\overline{A}w = A\overline{A}Ax - \sigma A\overline{A}x = A(\lambda\overline{x}) - \sigma\lambda x = \lambda w.$$

这说明存在 $\alpha \in \mathbb{R}$, $w = \alpha x$,从而 $A\overline{x} - \sigma x = \alpha x$.

再看 $A\overline{x}$ 与 x 线性无关(当 λ 是 $A\overline{A}$ 的多重特征值时可能出现此种情况).此时,向量

$$y = A\overline{x} + \mu x \neq 0, \quad \forall \mu \in \mathbb{C}.$$

选取 $\mu \in \mathbb{C}$,使得 $|\mu|^2 = \mu\overline{\mu} = \lambda$,则有

$$A\overline{y} = A\left(\overline{A}x + \overline{\mu x}\right) = A\overline{A}x + \overline{\mu}A\overline{x} = \lambda x + \overline{\mu}A\overline{x}$$
$$= \mu\overline{\mu}x + \overline{\mu}A\overline{x} = \overline{\mu}\left(A\overline{x} + \mu x\right) = \overline{\mu}y.$$

因此,无论 $A\overline{x}$ 与 x 线性相关还是线性无关,必存在非零向量 $v \in \mathbb{C}^n$ 和 $a \in \mathbb{C}$,使得

$$A\overline{v} = av. \tag{7.2}$$

此恒等式当 v 乘以正数时不会改变,故可假定 v 是单位向量,即 $\|v\|_2 = 1$.注意到

$$e^{-i\theta}A\overline{v} = A\overline{(e^{i\theta}v)} = e^{-i\theta}av = \left(e^{-i2\theta}a\right)\left(e^{i\theta}v\right), \quad \forall \theta \in \mathbb{R}, \tag{7.3}$$

$e^{i\theta}v$ 也是单位向量;并选取 $\theta \in \mathbb{R}$ 使得 $e^{-i2\theta}a \geq 0$.从(7.3)推出,存在

单位向量,仍记作v,使得

$$A\bar{v}=\sigma v,\quad \sigma\equiv \mathrm{e}^{-\mathrm{i}2\theta}a=+\sqrt{\lambda}\geq 0. \tag{7.4}$$

现在,从(7.4)的v出发,扩充成$\mathbb{C}^n$中的一组正交基

$$\left\{v,v^{(2)},\cdots,v^{(n)}\right\}.$$

令

$$V_1\equiv\begin{bmatrix}v & v^{(2)} & \cdots & v^{(n)}\end{bmatrix}\in\mathbb{C}^{n\times n},$$

这是一个酉矩阵.$\bar{V}_1^{\mathrm{T}}A\bar{V}_1$的第1列元素为

$$\left(v^{(i)}\right)^* A\bar{v}=\sigma\left(v^{(i)}\right)^* v=\sigma\delta_{i1},\quad i=1,\cdots,n,$$

其中$v^{(1)}\equiv v$,δ_{ij}是 Kronecker 记号.从而$\bar{V}_1^{\mathrm{T}}A\bar{V}_1$有如下分块形式:

$$\bar{V}_1^{\mathrm{T}}A\bar{V}_1=\begin{bmatrix}\sigma & w^{\mathrm{T}}\\ 0 & A_2\end{bmatrix},\ w\in\mathbb{C}^{n-1},A_2\in\mathbb{C}^{(n-1)\times(n-1)}. \tag{7.5}$$

由此,

$$\begin{aligned}V_1^*A\bar{A}V_1&=(\bar{V}_1^{\mathrm{T}}A\bar{V}_1)\overline{(\bar{V}_1^{\mathrm{T}}A\bar{V}_1)}\\&=\begin{bmatrix}\sigma^2 & \sigma\bar{w}^{\mathrm{T}}+w^{\mathrm{T}}\overline{A_2}\\ 0 & A_2\overline{A_2}\end{bmatrix}.\end{aligned}$$

表明$A\bar{A}$的特征值(依假设全为非负的)是σ^2及$A_2\overline{A_2}$的特征值.因而经以上约化过程而得的矩阵$A_2\in\mathbb{C}^{(n-1)\times(n-1)}$也具有$A_2\overline{A_2}$的所有特征值非负的性质.

这样,对A_2及其后继可以重复以上约化过程,总共至多$n-1$次,最后得

$$\bar{V}_{n-1}^{\mathrm{T}}\cdots\bar{V}_1^{\mathrm{T}}A\bar{V}_1\cdots\bar{V}_{n-1}=\begin{bmatrix}\sigma_1 & & \bigstar\\ & \ddots & \\ 0 & & \sigma_n\end{bmatrix}\equiv\Delta, \tag{7.6}$$

其中Δ是上三角矩阵,具有非负主对角元素$\sigma_1,\cdots,\sigma_n$.令

$$U \equiv V_1 \cdots V_{n-1},$$

从(7.6)即得(7.1). □

10.7.3 Takagi 分解定理 设 $A \in \mathbb{C}^{n\times n}$ 是对称矩阵,则存在酉矩阵 $U \in \mathbb{C}^{n\times n}$ 和非负对角矩阵 $\Sigma = \operatorname{diag}(\sigma_1, \cdots, \sigma_n)$,使得

$$A = U\Sigma U^{\mathrm{T}}, \tag{7.7}$$

U 的列是 $A\overline{A}$ 的特征向量正交组,Σ 的对应对角元素是 $A\overline{A}$ 的相应特征值的非负平方根.

证 因 $A = A^{\mathrm{T}}$,故

$$\overline{A} = A^*, A\overline{A} = AA^*.$$

如果 $x \neq 0$ 是自伴矩阵 AA^* 的任一特征向量,且 $AA^*x = \lambda x$,那么

$$x^*\lambda x = \lambda(x^*x) = x^*AA^*x = (A^*x)^*(A^*x). \tag{7.8}$$

由于

$$y^*y \geq 0, \quad \forall y \in \mathbb{C}^n,$$

$y^*y = 0$ 当且仅当 $y = 0$,从(7.8)得

$$\lambda = (A^*x)^*(A^*x)/x^*x \geq 0,$$

所以 $A\overline{A}$ 的所有特征值是非负的.依 **10.7.2**,存在酉矩阵 $U \in \mathbb{C}^{n\times n}$ 和上三角矩阵 $\Delta \in \mathbb{C}^{n\times n}$,

$$\Delta \equiv \begin{bmatrix} \sigma_1 & & \bigstar \\ & \ddots & \\ 0 & & \sigma_n \end{bmatrix}, \quad \sigma_1, \cdots, \sigma_n \geq 0,$$

使得 $A = U\Delta U^{\mathrm{T}}$.然而

$$U\Delta U^{\mathrm{T}} = A = A^{\mathrm{T}} = U\Delta^{\mathrm{T}}U^{\mathrm{T}},$$

推出

$$\Delta = \Delta^{\mathrm{T}} \equiv \Sigma$$

是对角矩阵.最后

$$A\overline{A}=U\Sigma U^{\mathrm{T}}\overline{U}\Sigma U^{*}=U\Sigma^{2}U^{*}$$

这是自伴矩阵 $A\overline{A}$ 的酉对角化，U 的列是 $A\overline{A}$ 的特征向量. □

10.7.4 推论 设 $A\in\mathbb{C}^{n\times n}$ 是对称矩阵，$A\overline{A}$ 的特征值是不同的，而且有酉对角化

$$A\overline{A}=V\Sigma^{2}V^{*},\quad \Sigma=\operatorname{diag}(\sigma_1,\cdots,\sigma_n),\sigma_1,\cdots,\sigma_n\geq 0, \tag{7.9}$$

则存在对角矩阵

$$D=\operatorname{diag}\left(\mathrm{e}^{\mathrm{i}\theta_1},\cdots,\mathrm{e}^{\mathrm{i}\theta_n}\right),\quad \theta_1,\cdots,\theta_n\in\mathbb{R}, \tag{7.10}$$

使得

$$A=U\Sigma U^{\mathrm{T}},\qquad U\equiv VD. \tag{7.11}$$

D 相应于 Σ 的非零对角元素的对角元素是由关系式

$$A\overline{V}=V\Sigma D^{2}$$

确定的；D 相应于 Σ 的零对角元素的对角元素可取作 1.

证 设 λ 是 $A\overline{A}$ 的特征值，$x\neq 0$ 是对应的特征向量.依假设，λ 是单特征值，从 10.7.2 的证明得知，$A\overline{x}$ 必与 x 线性相关，有

$$A\overline{x}=ax,\ \ a=\sigma\mathrm{e}^{\mathrm{i}2\theta},\ \ \sigma=+\sqrt{\lambda},\ \ \theta\in\mathbb{R}$$

和

$$A\overline{A}x=\sigma^{2}x.$$

由此，在条件(7.9)下，必有

$$\begin{aligned}A\overline{V}&\equiv A\left[\overline{x^{(1)}}\quad\cdots\quad\overline{x^{(n)}}\right]\\&=\left[x^{(1)}\quad\cdots\quad x^{(n)}\right]\operatorname{diag}\left(\sigma_1\mathrm{e}^{\mathrm{i}2\theta_1},\cdots\sigma_n\mathrm{e}^{\mathrm{i}2\theta_n}\right)\\&=V\Sigma D^{2}.\end{aligned}$$

最后，

$$\begin{aligned}A&=A\overline{V}V^{\mathrm{T}}=V\Sigma D^{2}V^{\mathrm{T}}\\&=(VD)\Sigma(VD)^{\mathrm{T}}=U\Sigma U^{\mathrm{T}}.\end{aligned}$$

□

10.7.5 推论 设 $A\in\mathbb{C}^{n\times n}$，$A$ 对称的充分必要条件是存在矩阵

$S \in \mathbb{C}^{n\times n}$ 使得

$$A = SS^{\mathrm{T}}. \tag{7.12}$$

一种选取是

$$S = UD,$$

其中 U 是酉矩阵,而

$$D = \operatorname{diag}\left(\sqrt{\sigma_1}, \cdots, \sqrt{\sigma_n}\right),$$

$\sigma_1, \cdots, \sigma_n$ 是 A 的奇异值.此时, $\operatorname{rank} S = \operatorname{rank} A$.

证 若 A 对称,则依 **10.7.3**,

$$A = U\Sigma U^{\mathrm{T}} = \left(U\Sigma^{1/2}\right)\left(U\Sigma^{1/2}\right)^{\mathrm{T}} \equiv SS^{\mathrm{T}},$$

其中 $\Sigma^{1/2} = \operatorname{diag}\left(\sqrt{\sigma_1}, \cdots, \sqrt{\sigma_n}\right) \equiv D$.

反之,若 $A = SS^{\mathrm{T}}$,则 A 显然对称. □

10.7.6 定理 设 $A \in \mathbb{C}^{n\times n}$, A 对称且正规的充分必要条件是存在正交矩阵 $Q \in \mathbb{R}^{n\times n}$ 和对角矩阵 $\Lambda \in \mathbb{C}^{n\times n}$ 使得

$$A = Q\Lambda Q^{\mathrm{T}}. \tag{7.13}$$

证 先证必要性.设 $A = B + \mathrm{i}C \in \mathbb{C}^{n\times n}$, $B, C \in \mathbb{R}^{n\times n}$,则 A 对称等价于 B 和 C 同为实对称矩阵.又从 A 正规,有

$$\begin{aligned} AA^* &= \left(B^2 + C^2\right) + \mathrm{i}(CB - BC) \\ &= \left(B^2 + C^2\right) + \mathrm{i}(BC - CB) = A^*A, \end{aligned}$$

推出 B 和 C 可以交换.如此,并依 **5.1.9**,存在正交矩阵 $Q \in \mathbb{R}^{n\times n}$,使得 B 和 C 同时对角化:

$$B = QD_1Q^{\mathrm{T}}, \qquad C = QD_2Q^{\mathrm{T}},$$

其中 $D_1, D_2 \in \mathbb{R}^{n\times n}$ 是对角矩阵.因此

$$\begin{aligned} A = B + \mathrm{i}C &= QD_1Q^{\mathrm{T}} + \mathrm{i}QD_2Q^{\mathrm{T}} \\ &= Q\left(D_1 + \mathrm{i}D_2\right)Q^{\mathrm{T}} = Q\Lambda Q^{\mathrm{T}}, \end{aligned}$$

这里 $\Lambda \equiv D_1 + \mathrm{i}D_2$ 是对角矩阵.

再证充分性.设(7.13)成立,其中 $Q \in \mathbb{R}^{n\times n}$ 是正交矩阵,$\Lambda \in \mathbb{C}^{n\times n}$ 是对角矩阵,则 $\Lambda = \Lambda^{\mathrm{T}}$,而且

$$AA^* = Q\Lambda Q^{\mathrm{T}} Q\overline{\Lambda} Q^{\mathrm{T}} = Q|\Lambda|^2 Q^{\mathrm{T}}$$
$$= Q\overline{\Lambda} Q^{\mathrm{T}} Q\Lambda Q^{\mathrm{T}} = A^* A,$$

所以 A 对称且正规. □

例 对称正规复矩阵的一个有用而简单的例子是

$$S = \frac{1}{\sqrt{2}}(I + \mathrm{i}E), \tag{7.14}$$

其中 E 是反序单位矩阵

$$E = \begin{bmatrix} 0 & & 1 \\ & \cdot^{\cdot^{\cdot}} & \\ 1 & & 0 \end{bmatrix}.$$

事实上,$E^2 = I$,

$$S\overline{S} = \frac{1}{2}(I + \mathrm{i}E)(I - \mathrm{i}E) = \frac{1}{2}(I - \mathrm{i}E + \mathrm{i}E + I) = I.$$

表明 S 是对称的酉矩阵.

10.7.7 定理 每个矩阵 $A \in \mathbb{C}^{n\times n}$ 相似于一个对称矩阵.

证 设 $A \in \mathbb{C}^{n\times n}$,将其 Jordan 标准形 J 表示为 Jordan 块的直和:

$$J = J_{n_1}(\lambda_1) \oplus \cdots \oplus J_{n_k}(\lambda_k), \tag{7.15}$$

其中 Jordan 块

$$J_{n_j}(\lambda_j) \in \mathbb{C}^{n_j \times n_j}, \ \lambda_j = \alpha_j + \mathrm{i}\beta_j, \quad j = 1, \cdots, k,$$

而且 $n_1 + \cdots + n_k = n$.设

$$S_{n_j} \in \mathbb{C}^{n_j \times n_j}, \quad j = 1, \cdots, k,$$

当 $n_j \geq 2$ 时,$S_{n_j} = \dfrac{1}{\sqrt{2}}(I + \mathrm{i}E)$ 是形如(7.14)的矩阵;当 $n_j = 1$ 时,$S_1 \equiv [1]$.令

$$T = S_{n_1} \oplus \cdots \oplus S_{n_k},$$

则成立

$$\begin{aligned} TJT^{-1} &= TJ\overline{T} \\ &= \left(S_{n_1} J_{n_1}(\lambda_1)\overline{S_{n_1}}\right) \oplus \cdots \oplus \left(S_{n_k} J_{n_k}(\lambda_k)\overline{S_{n_k}}\right). \end{aligned} \tag{7.16}$$

其中,当 $n_j = 1$ 时, $J_1(\lambda_j) = [\lambda_j]$,

$$S_{n_j}(\lambda_j) = S_1(\lambda_j) \equiv S_1 J_1(\lambda_j)\overline{S_1} = [\lambda_j] = [\alpha_j + \mathrm{i}\beta]. \tag{7.17}$$

当 $n_j \geq 2$ 时, $J_{n_j}(\lambda_j) = \lambda_j I_{n_j} + N_{n_j}$,

$$N_{n_j} = \begin{bmatrix} 0 & 1 & & 0 \\ & \ddots & \ddots & \\ & & \ddots & 1 \\ 0 & & & 0 \end{bmatrix} \in \mathbb{R}^{n_j \times n_j}.$$

于是

$$\begin{aligned} S_{n_j}(\lambda_j) &\equiv S_{n_j} J_{n_j}(\lambda_j)\overline{S_{n_j}} = \lambda_j S_{n_j} I_{n_j} \overline{S_{n_j}} + S_{n_j} N_{n_j} \overline{S_{n_j}} \\ &= \lambda_j I_{n_j} + S_{n_j} N_{n_j} \overline{S_{n_j}}. \end{aligned} \tag{7.18}$$

注意到 $S_{n_j} = \dfrac{1}{\sqrt{2}}\left(I_{n_j} + \mathrm{i}E_{n_j}\right)$,及

$$E_{n_j} N_{n_j} E_{n_j} = \begin{bmatrix} 0 & & 0 \\ 1 & \ddots & \\ & \ddots & \ddots \\ 0 & & 1 & 0 \end{bmatrix},$$

$$E_{n_j} N_{n_j} = \begin{bmatrix} 0 & & 0 \\ & \cdot^{\cdot^{\cdot}} & 1 \\ & \cdot^{\cdot^{\cdot}} & \cdot^{\cdot^{\cdot}} \\ 0 & 1 & 0 \end{bmatrix}, \quad N_{n_j} E_{n_j} = \begin{bmatrix} 0 & 1 & 0 \\ & \cdot^{\cdot^{\cdot}} & \cdot^{\cdot^{\cdot}} \\ 1 & \cdot^{\cdot^{\cdot}} & \\ 0 & & 0 \end{bmatrix}.$$

N_{n_j} 酉相似于矩阵

$$S_{n_j}N_{n_j}S_{n_j}^{-1}=S_{n_j}N_{n_j}\overline{S_{n_j}}=\frac{1}{2}\left(I_{n_j}+\mathrm{i}E_{n_j}\right)N_{n_j}\left(I_{n_j}-\mathrm{i}E_{n_j}\right)$$

$$=\frac{1}{2}\left(N_{n_j}+E_{n_j}N_{n_j}E_{n_j}\right)+\frac{\mathrm{i}}{2}\left(E_{n_j}N_{n_j}-N_{n_j}E_{n_j}\right)$$

$$=\frac{1}{2}\begin{bmatrix}0&1&&0\\1&\ddots&\ddots&\\&\ddots&\ddots&1\\0&&1&0\end{bmatrix}+\frac{\mathrm{i}}{2}\begin{bmatrix}0&&-1&0\\&\cdot^{\cdot^{\cdot}}&\cdot^{\cdot^{\cdot}}&1\\-1&\cdot^{\cdot^{\cdot}}&\cdot^{\cdot^{\cdot}}&\\0&1&&0\end{bmatrix}.$$

由此及(7.18),当 $n_j\geq 2$ 时,

$$S_{n_j}\left(\lambda_j\right)\equiv S_{n_j}J_{n_j}\left(\lambda_j\right)\overline{S_{n_j}}=\frac{1}{2}\begin{bmatrix}2\alpha_j&1&&0\\1&\ddots&\ddots&\\&\ddots&\ddots&1\\0&&1&2\alpha_j\end{bmatrix}$$

$$+\frac{\mathrm{i}}{2}\left\{\begin{bmatrix}2\beta_j&&0\\&\ddots&\\0&&2\beta_j\end{bmatrix}+\begin{bmatrix}0&&-1&0\\&\cdot^{\cdot^{\cdot}}&\cdot^{\cdot^{\cdot}}&1\\-1&\cdot^{\cdot^{\cdot}}&\cdot^{\cdot^{\cdot}}&\\0&1&&0\end{bmatrix}\right\}.$$

这显然是对称矩阵.

最后, A 相似于对称矩阵

$$TJT^{-1}=TJ\overline{T}=S_{n_1}\left(\lambda_1\right)\oplus\cdots\oplus S_{n_k}\left(\lambda_k\right).$$ □

10.7.8 推论 设 $A\in\mathbb{C}^{n\times n}$,存在非奇异矩阵 $S\in\mathbb{C}^{n\times n}$ 和酉矩阵 $U\in\mathbb{C}^{n\times n}$,使得

$$(US)A\left(\overline{U}S\right)^{-1}$$

是非负对角矩阵.

证　依 **10.7.7**,存在非奇异矩阵 $S\in\mathbb{C}^{n\times n}$,使得 SAS^{-1} 是对称矩阵;再依 **10.7.3**,存在酉矩阵 $U\in\mathbb{C}^{n\times n}$,使得

$$U\left(SAS^{-1}\right)U^{\mathrm{T}}\equiv(US)A\left(\bar{U}S\right)^{-1}$$

是非负对角矩阵. □

10.7.9 推论　设 $A\in\mathbb{C}^{n\times n}$.存在对称矩阵 $B,C\in\mathbb{C}^{n\times n}$,使得

$$A=BC.$$

而且可以选取 B 或 C 是非奇异矩阵.

证　依 **10.7.7**,存在表示式 $A=SFS^{-1}$,其中 $F=F^{\mathrm{T}}$,而 S 是非奇异矩阵.因此

$$A=\left(SFS^{\mathrm{T}}\right)\left(S^{\mathrm{T}}\right)^{-1}S^{-1}=\left(SFS^{\mathrm{T}}\right)\left(SS^{\mathrm{T}}\right)^{-1}=BC,$$

其中 $B\equiv SFS^{\mathrm{T}}$ 和 $C\equiv\left(SS^{\mathrm{T}}\right)^{-1}$ 是对称矩阵,C 是非奇异的.又

$$A=\left(SS^{\mathrm{T}}\right)\left(S^{-1}\right)^{\mathrm{T}}FS^{-1},$$

此时 $B\equiv SS^{\mathrm{T}}$ 和 $C\equiv\left(S^{-1}\right)^{\mathrm{T}}FS^{-1}$ 对称,B 非奇异. □

10.7.10 引理　设 $x^{(1)},\cdots,x^{(k)}\in\mathbb{C}^{n}$,$k\le n$,则存在

$$y^{(1)},\cdots,y^{(k)}\in\mathbb{C}^{n},$$

满足

$$\operatorname{span}\left\{x^{(1)},\cdots,x^{(k)}\right\}=\operatorname{span}\left\{y^{(1)},\cdots,y^{(k)}\right\}\tag{7.19}$$

和

$$\begin{aligned}&y_i^{\mathrm{T}}y_j=0,\quad i,j=1,\cdots,k,\\&y_i^{\mathrm{T}}y_i=\begin{cases}1,&i=1,\cdots,r,\\0,&i=r+1,\cdots,k,\end{cases}\end{aligned}\tag{7.20}$$

其中

$$r=\operatorname{rank}X^{\mathrm{T}}X,\quad X\equiv\left[x^{(1)}\quad\cdots\quad x^{(k)}\right]\in\mathbb{C}^{n\times k}.$$

证　因为 $X^{\mathrm{T}}X\in\mathbb{C}^{k\times k}$ 是对称矩阵,依 **10.7.3**,存在酉矩阵

$U \in \mathbb{C}^{k\times k}$ 和 $\Sigma = \mathrm{diag}(\sigma_1,\cdots,\sigma_k)$,

$$\sigma_1 \geq \cdots \geq \sigma_r > \sigma_{r+1} = \cdots = \sigma_k = 0, \quad \mathrm{rank}X^{\mathrm{T}}X = r,$$

使得

$$X^{\mathrm{T}}X = U\Sigma U^{\mathrm{T}}.$$

令

$$D = \mathrm{diag}\left(\sqrt{\sigma_1},\cdots,\sqrt{\sigma_r},1,\cdots,1\right) \in \mathbb{R}^{k\times k}$$

和

$$E_r = \mathrm{diag}\{\underbrace{1,\cdots,1}_{r},0,\cdots,0\} \in \mathbb{R}^{k\times k},$$

则有

$$X^{\mathrm{T}}X = (UD)E_r(UD)^{\mathrm{T}} = S^{\mathrm{T}}E_rS,$$

其中 $S \equiv DU^{\mathrm{T}}$ 非奇异.于是

$$\left(XS^{-1}\right)^{\mathrm{T}}\left(XS^{-1}\right) = E_r.$$

因此,若令

$$XS^{-1} \equiv Y = \begin{bmatrix} y^{(1)} & \cdots & y^{(k)} \end{bmatrix} \in \mathbb{C}^{n\times k},$$

则 $Y^{\mathrm{T}}Y = E_r$,向量 $y^{(1)},\cdots,y^{(k)}$ 满足(7.19)和(7.20). □

10.7.11 定理 设 $A \in \mathbb{C}^{n\times n}$ 是对称矩阵, A 可对角化的充分必要条件是 A 为复正交可对角化.也就是说,存在对角矩阵 $\Lambda \in \mathbb{C}^{n\times n}$ 和非奇异矩阵 $S \in \mathbb{C}^{n\times n}$ 使得

$$A = S\Lambda S^{-1}$$

的充分必要条件是

$$A = Q\Lambda Q^{\mathrm{T}},$$

其中 $Q \in \mathbb{C}^{n\times n}$ 满足

$$Q^{\mathrm{T}}Q = I.$$

证 充分性显然.下面证明必要性.

依假设 $A = A^{\mathrm{T}}$.设 $x, y \in \mathbb{C}^n$ 是 A 的特征向量,

$$Ax = \lambda x, \qquad Ay = \mu y.$$

若$\lambda \neq \mu$,则有

$$y^{\mathrm{T}} Ax = y^{\mathrm{T}} \lambda x = \lambda y^{\mathrm{T}} x$$

和

$$y^{\mathrm{T}} Ax = (Ay)^{\mathrm{T}} x = (\mu y)^{\mathrm{T}} x = \mu y^{\mathrm{T}} x.$$

因此

$$\lambda y^{\mathrm{T}} x = \mu y^{\mathrm{T}} x,$$

推出$y^{\mathrm{T}} x = 0$.

如果A可对角化,

$$A = S \Lambda S^{-1},$$

假定

$$\Lambda = \Lambda_1 \oplus \cdots \oplus \Lambda_d, \qquad \Lambda_i \equiv \lambda_i I_{n_i}, i = 1, \cdots, d,$$

其中

$$n_1 + \cdots + n_d = n, \quad \lambda_i \neq \lambda_j \text{ 当 } i \neq j$$

相应地将S沿列向分块

$$S = [S_1 \quad \cdots \quad S_d], \qquad S_i \in \mathbb{C}^{n \times n_i}, i = 1, \cdots, d.$$

注意到

$$S_i^{\mathrm{T}} S_j = 0 \in \mathbb{C}^{n_i \times n_j}, \quad \text{当 } i \neq j,$$

而且$S_i^{\mathrm{T}} S_i \in \mathbb{C}^{n_i \times n_i}$是非奇异的,$i = 1, \cdots, d$.依 **10.7.10**,对于$S_i$,存在非奇异矩阵$R_i \in \mathbb{C}^{n_i \times n_i}$,使得$Q_i \equiv S_i R_i$满足

$$Q_i^{\mathrm{T}} Q_i = R_i^{\mathrm{T}} S_i^{\mathrm{T}} S_i R_i = I_{n_i}.$$

因为有

$$Q_i^{\mathrm{T}} Q_j = R_i^{\mathrm{T}} S_i^{\mathrm{T}} S_j R_j = 0, \quad \text{当 } i \neq j,$$

和

$$AQ_i = AS_i R_i = \lambda_i S_i R_i = \lambda_i Q_i, i = 1, \cdots, d,$$

所以

$$Q \equiv \begin{bmatrix} Q_1 & \cdots & Q_d \end{bmatrix} \in \mathbb{C}^{n\times n}$$

是复正交矩阵,而且 $A = Q\Lambda Q^{\mathrm{T}}$. □

10.7.12 定义 设 $A \in \mathbb{C}^{n\times n}$ 是三对角矩阵

$$A = \begin{bmatrix} \alpha_1 & \beta_2 & & 0 \\ \gamma_2 & \alpha_2 & \ddots & \\ & \ddots & \ddots & \beta_n \\ 0 & & \gamma_n & \alpha_n \end{bmatrix}, \tag{7.21}$$

如果满足

$$\alpha_i \in \mathbb{R}, \quad i = 1, \cdots, n \tag{7.22}$$

和

$$\beta_i \gamma_i > 0, \quad i = 2, \cdots, n, \tag{7.23}$$

则称 A 为**拟对称三对角矩阵**(quasi-symmetric tri-diagonal matrix).

10.7.13 定理 设 $A \in \mathbb{C}^{n\times n}$ 是拟对称三对角矩阵,则 A 相似于实对称三对角矩阵.

证 设 A 形如(7.21).定义

$$D = \mathrm{diag}(d_{11}, \cdots, d_{nn}),$$

其中

$$d_{11} = 1; \quad d_{ii} = \sqrt{\frac{\gamma_2 \cdots \gamma_i}{\beta_2 \cdots \beta_i}}, \; i = 2, \cdots, n,$$

则有

$$D^{-1} A D = T,$$

其中 T 是实对称三对角矩阵,

$$t_{ii} = \alpha_i, \; i = 1, \cdots, n;$$

$$t_{i,i+1} = t_{i+1,i} = \sqrt{\beta_{i+1}\gamma_{i+1}}, \; i = 1, \cdots, n-1.$$ □

10.8 辛矩阵

10.8.1 引言 实 Euclid 空间$\mathbb{R}^n$的内积是**对称**双线性函数.

在 **1.7.26** 中是把具有**保内积**性质的矩阵定义为正交矩阵.具体地说,$O\in\mathbb{R}^{n\times n}$是正交的,如果

$$(Ox,Oy)=(x,y),\quad \forall x,y\in\mathbb{R}^n.$$

现在考察**保**形如(x,Ay)的非奇异双线性交替函数的矩阵;当$A\in\mathbb{R}^{n\times n}$是斜自伴矩阵且$\det A\neq 0$时,称$(x,Ay)$为**非奇异双线性交替函数**.

"交替"性质决定矩阵$(-\lambda)^m\det(\lambda I-J)=\lambda^m\left(-\left(1+\lambda^2\right)\right)^m$的阶必须是偶数,$n=2m$;特别地,只须取$A=J$:

$$J\equiv\begin{bmatrix}0 & I\\ -I & 0\end{bmatrix}\tag{8.1}$$

其中I是m阶单位矩阵.

事实上,J是斜自伴矩阵.依 **3.4.2**,斜自伴矩阵所有特征值为纯虚数,而且存在特征向量正交基.

可以证明(留作练习):任何非奇异的实$2m$阶斜自伴矩阵A可以表示成

$$A=FJF^{\mathrm{T}}\tag{8.2}$$

其中F是某个实矩阵,$\det F\neq 0$.

10.8.2 定理 由(8.1)定义的$2m$阶矩阵J有如下性质:

(1) $J^2=-I$,$J^{-1}=-J=J^{\mathrm{T}}$.

(2) 纯虚数i和$-\mathrm{i}$各为m重特征值.

(3) $\det J=1$.

证 (1)显然.

由于

$$\lambda I-J=\begin{bmatrix}\lambda I & -I\\ I & \lambda I\end{bmatrix},$$

而且

$$\begin{bmatrix}I & 0\\ I & -\lambda I\end{bmatrix}\begin{bmatrix}\lambda I & -I\\ I & \lambda I\end{bmatrix}=\begin{bmatrix}\lambda I & -I\\ 0 & -\left(1+\lambda^2\right)I\end{bmatrix},$$

两边取行列式,得对任何数 λ 成立

$$(-\lambda)^m\det(\lambda I-J)=\lambda^m\left(-\left(1+\lambda^2\right)\right)^m.$$

于是 J 的特征多项式为

$$\det(\lambda I-J)=\left(1+\lambda^2\right)^m,$$

由此推出(2).

现在根据(2),

$$\det J=\mathrm{i}^m(-\mathrm{i})^m=(-1)^m\mathrm{i}^{2m}=1.$$

(3)得证. □

10.8.3 定理 设 $J\in\mathbb{R}^{2m\times 2m}$ 是由(8.1)定义的矩阵.如果矩阵 $S\in\mathbb{R}^{2m\times 2m}$ 保 (x,Jy):

$$(Sx,JSy)=(x,Jy),\quad\forall x,y\in\mathbb{R}^{2m},\tag{8.3}$$

那么

$$S^{\mathrm{T}}JS=J.\tag{8.4}$$

反之亦然.

证 利用

$$(Sx,JSy)=\left(x,S^{\mathrm{T}}JSy\right).$$

如果(8.3)成立,那么

$$S^{\mathrm{T}}JSy=Jy,\quad\forall y\in\mathbb{R}^{2m},$$

由此推出(8.4).反过来从(8.4)得(8.3)更是显然的. □

10.8.4 定义 保双线性交替函数 (x,Jy) 的 $S\in\mathbb{R}^{2m\times 2m}$,等价地说,满足(8.4)的 $S\in\mathbb{R}^{2m\times 2m}$,称为**辛矩阵**(symplectic matrix).

$\mathbb{R}^{2m\times 2m}$ 中辛矩阵全体所成之集记作 $Sp(2m)$.

辛矩阵首先出现在受如下形式控制的 Hamilton 力学中:

$$\frac{\mathrm{d}}{\mathrm{d}t}u = JH_u, \tag{8.5}$$

其中 $J \in \mathbb{R}^{2m\times 2m}$ 由(8.1)定义, $u(t) \in \mathbb{R}^{2m}$, H 是 $\mathbb{R}^{2m}$ 上的光滑函数, H_u 是 H 的梯度.

下面引入一个有关概念及结论.

非线性映射 $u \mapsto v$ 称为**标准变换**,如果它的 Jacobi 矩阵

$$\left[\frac{\partial v_i}{\partial u_j}\right]$$

是辛矩阵.

标准变换将每个 Hamilton 方程(8.5)变换成如下另一个形式的 Hamilton 方程

$$\frac{\mathrm{d}}{\mathrm{d}t}v = JK_v,$$

其中 $K(v(u)) = H(u)$.

10.8.5 定理 (1) $Sp(2m)$在矩阵乘法下构成一个群.

(2) 若 S 是辛矩阵,则其转置 S^{T} 也是辛矩阵.

(3) 辛矩阵 S 必相似于其逆 S^{-1}.因此,若 λ 是 S 的特征值,则 λ^{-1} 也是 S 的特征值.

证 从(8.4)推出每个辛矩阵是可逆的.而且,若 S_1 和 S_2 是辛矩阵:

$$S_1^{\mathrm{T}} J S_1 = J, \qquad S_2^{\mathrm{T}} J S_2 = J,$$

则有

$$J = S_2^{\mathrm{T}} J S_2 = S_2^{\mathrm{T}} S_1^{\mathrm{T}} J S_1 S_2 = (S_1 S_2)^{\mathrm{T}} J (S_1 S_2).$$

推出 S_1S_2 也是辛矩阵,从而 $Sp(2m)$ 在矩阵乘法下构成一个群.

对(8.4)取逆,得

$$S^{-1}J\left(S^{\mathrm{T}}\right)^{-1}=J.$$

然后左乘 S,右乘 S^{T},有

$$J=SJS^{\mathrm{T}}=\left(S^{\mathrm{T}}\right)^{\mathrm{T}}J\left(S^{\mathrm{T}}\right).$$

表明 S^{T} 满足(8.4),即为辛矩阵.

对(8.4)右乘 S^{-1},左乘 J^{-1},得

$$J^{-1}S^{\mathrm{T}}J=S^{-1}.$$

表明 S^{-1} 相似于 S^{T}.又 S^{T} 相似于 S.所以 S 与 S^{-1} 相似. □

10.8.6 定理 设 S 是实变量 t 的可微函数,其值是辛矩阵.定义 $G(t)$ 为

$$\frac{\mathrm{d}}{\mathrm{d}t}S(t)=G(t)S(t), \tag{8.6}$$

则 G 形如

$$G(t)=JL(t), \qquad L\text{ 自伴}, \tag{8.7}$$

其中 J 由(8.1)定义.

反之,若 $S(t)$ 满足(8.6)和(8.7),而且 $S(0)$ 是辛矩阵,则 $S(t)$ 是一族辛矩阵.

证 因为对 t 的每一值,成立

$$\left(S(t)\right)^{\mathrm{T}}JS(t)=J.$$

对 t 求导:

$$\left(\frac{\mathrm{d}}{\mathrm{d}t}\left(S(t)\right)^{\mathrm{T}}\right)JS(t)+\left(S(t)\right)^{\mathrm{T}}J\frac{\mathrm{d}}{\mathrm{d}t}S(t)=0.$$

右乘 $\left(S(t)\right)^{-1}$,左乘 $\left(\left(S(t)\right)^{\mathrm{T}}\right)^{-1}$:

$$\left(\left(S(t)\right)^{\mathrm{T}}\right)^{-1}\left(\frac{\mathrm{d}}{\mathrm{d}t}\left(S(t)\right)^{\mathrm{T}}\right)J+J\left(\frac{\mathrm{d}}{\mathrm{d}t}S(t)\right)\left(S(t)\right)^{-1}=0. \tag{8.8}$$

利用(8.6)定义的G：

$$G(t)=\left(\frac{\mathrm{d}}{\mathrm{d}t}S(t)\right)(S(t))^{-1},$$

取其转置,

$$(G(t))^{\mathrm{T}}=\left((S(t))^{\mathrm{T}}\right)^{-1}\frac{\mathrm{d}}{\mathrm{d}t}(S(t))^{\mathrm{T}}.$$

将它们代入(8.8),得

$$(G(t))^{\mathrm{T}}J+JG(t)=0.$$

由此及 **10.8.2** 的(1),推出(8.7).

逆定理的证明留作练习. □

10.8.7 定理 对于辛矩阵S,$\lambda=1$或-1不可能是单重特征值.

证 用反证法.不妨设$\lambda=-1$是S的单重特征值,其特征向量为x：

$$Sx=-x. \tag{8.9}$$

两边左乘$S^{\mathrm{T}}J$并利用(8.4),得

$$Jx=-S^{\mathrm{T}}Jx. \tag{8.10}$$

表明Jx是S^{T}的相应特征值-1的特征向量.

任取一正定矩阵L,令$G=JL$.定义单参数矩阵族:

$$S(t)\equiv \mathrm{e}^{tG}S,$$

它满足

$$\frac{\mathrm{d}}{\mathrm{d}t}S(t)=GS(t),\quad S(0)=S. \tag{8.11}$$

依 **10.8.6**,对所有t,$S(t)$是辛矩阵.因为已假定$S(0)$以-1为单重特征值,因此对于足够小的t,$S(t)$有一近乎-1的单重特征值.然而$S(t)$的该特征值λ必须等于-1.不然的话,因$(S(t))^{-1}$和$S(t)$相似,λ^{-1}将是$S(t)$又一个近乎-1的特征值,与 **9.5.14** 矛盾.依 **9.5.16**,对应的特征向量x是t的可微函数.对

$$S(t)x(t)=-x(t) \tag{8.12}$$

求导,得

$$\left(\frac{\mathrm{d}}{\mathrm{d}t}S(t)\right)x(t)+S(t)\frac{\mathrm{d}}{\mathrm{d}t}x(t)=-\frac{\mathrm{d}}{\mathrm{d}t}x(t).$$

再利用(8.11)及(8.12),有

$$Gx(t)=\frac{\mathrm{d}}{\mathrm{d}t}x(t)+S(t)\frac{\mathrm{d}}{\mathrm{d}t}x(t).$$

由此及(8.10),

$$\begin{aligned}\left(Gx(t),Jx(t)\right)&=\left(\frac{\mathrm{d}}{\mathrm{d}t}x(t),Jx(t)\right)+\left(S(t)\frac{\mathrm{d}}{\mathrm{d}t}x(t),Jx(t)\right)\\&=\left(\frac{\mathrm{d}}{\mathrm{d}t}x(t),Jx(t)\right)+\left(\frac{\mathrm{d}}{\mathrm{d}t}x(t),\left(S(t)\right)^{\mathrm{T}}Jx(t)\right)\\&=\left(\frac{\mathrm{d}}{\mathrm{d}t}x(t),Jx(t)\right)-\left(\frac{\mathrm{d}}{\mathrm{d}t}x(t),Jx(t)\right)=0.\end{aligned}$$

代入$G=JL$,

$$\left(JLx(t),Jx(t)\right)=\left(Lx(t),J^{\mathrm{T}}Jx(t)\right)=\left(Lx(t),x(t)\right)=0,$$

因为L是正定的,推出$x(t)=0$.这样,得出了矛盾. □

可以进一步证明:辛矩阵不可能以1或-1作为奇重数的特征值.

现在可以概括一下辛矩阵S的谱.以上已讨论S的特征值λ为1或-1(如果有的话)的情形,姑且不再考虑.因为S是实的,它的复特征值必成对出现,也就是说,若λ是特征值,则$\bar{\lambda}$也是特征值.又依**10.8.5**的(3),若λ是S的特征值,则λ^{-1}也是特征值.因此,辛矩阵S的特征值按四个一组出现:

$$\lambda,\ \bar{\lambda},\ \lambda^{-1},\ \overline{\lambda^{-1}}.$$

除去两种例外:

(1) λ在单位圆上,$|\lambda|=1$,此时$\lambda^{-1}=\bar{\lambda}$,故只有两个成一组:$\lambda,\bar{\lambda}$.

(2) λ 是实的,$\bar{\lambda}=\lambda$,也只有两个成一组:λ,λ^{-1}.

10.8.8 定理 辛矩阵 S 的行列式等于1.

证 对(8.4)取行列式得

$$(\det S)^2=1.$$

因此只须排除 $\det S$ 为负的可能性.矩阵的行列式等于其所有特征值之积.S 的复特征值成对出现,其积是正的.S 的不等于1或-1的实特征值是 λ 和 λ^{-1} 成对出现,其积是正的;而且如以-1为特征值,其重数只能是偶数,其积也是正的. □

10.9 整数矩阵和幺模矩阵

10.9.1 定义 $A\in\mathbb{R}^{m\times n}$ 称为**整数矩阵**(integer matrix),如果其所有元素都是整数.

整数矩阵在线性规划,格论和控制论等学科中有重要应用.例如概括地说,线性规划模型是在线性方程组

$$Ax=b \tag{9.1}$$

约束下求某个线性函数

$$\sum c_i x_i$$

的最大值(或最小值),这里 $A\in\mathbb{R}^{m\times n}$,$b\in\mathbb{R}^m$,$x=(x_1,\cdots,x_n)^{\mathrm{T}}$ 是待求向量.因为 $x_1,\cdots,x_n$ 代表物理量,它们常是非负的;而且,在许多情况中,A 和 b 的所有元素全为整数,解 x 也必须取整数分量(这构成**整数规划**问题).

整数矩阵和多项式矩阵(λ-矩阵)有非常类似的性质,原因是多项式全体和整数全体两者在各自的加法与乘法运算下均构成

一个环.

10.9.2 定理 设 $A \in \mathbb{R}^{m\times n}$ 是整数矩阵, $\mathrm{rank}A = r$, 则存在非奇异整数矩阵 $P \in \mathbb{R}^{m\times m}$ 和 $Q \in \mathbb{R}^{n\times n}$, 使得

$$PAQ = D \equiv \begin{bmatrix} \hat{D} & 0 \\ 0 & 0 \end{bmatrix}, \quad \hat{D} = \mathrm{diag}(a_1, \cdots, a_r), \tag{9.2}$$

其中 $a_1, \cdots, a_r$ 是整数, a_i 整除 a_{i+1}, $i = 1, \cdots, r-1$, 而且

$$\det P = \pm 1, \quad \det Q = \pm 1. \tag{9.3}$$

这是与 λ-矩阵的 **9.6.9** 相平行的结论.不难仿照 **9.6.9** 加以证明,留作练习.

(9.2)中的 D 称为整数矩阵 A 的 **Smith 标准形**.

10.9.3 定理 设 $A \in \mathbb{R}^{m\times n}$ 是整数矩阵, $m \le n$. 方程组(9.1)对所有整数向量 b 有整数解的充分必要条件是 A 的 Smith 标准形

$$D = [I_m \quad 0].$$

定理证明从略.

10.9.4 定义 如果 $A \in \mathbb{R}^{m\times n}$ 是整数矩阵,

$$r \equiv \mathrm{rank}A = \min\{m, n\},$$

而且 A 的所有非零 $r \times r$ 子式等于 1 或 -1, 则称 A 为**幺模矩阵**(unimodular matrix).

如果 A 是幺模矩阵,而且还有其各阶子式均等于 0, 1 或 -1, 则称 A 为**全幺模矩阵**(totally unimodular matrix).

特别,当 $m = n$ 时,整数矩阵 A 是幺模矩阵,如果

$$\det A = 1 \quad 或 \quad \det A = -1.$$

显然:

(1) 全幺模矩阵的所有元素均为 $0,1$ 或 -1.

(2) 两个 $n\times n$ 的幺模矩阵之积仍是幺模矩阵;在矩阵乘法下,$n\times n$ 的幺模矩阵全体构成一个群.

10.9.5 定义 设 $m\le n$, $x=\left(x_1,\cdots,x_n\right)^{\mathrm{T}}$ 是方程组(9.1)的解,而且 x 有 m 个分量不为零,则称 x 是(9.1)的**基本解**.

在整数规划问题中要求基本整数解.

10.9.6 定理 设 $m\le n$.对于任何整数向量 b,方程组(9.1)的所有基本解为整数向量的充分必要条件是 A 为幺模矩阵.

定理证明从略.

10.9.7 定理 设 $A\in\mathbb{R}^{m\times n}$ 是整数矩阵.如果对于任何整数向量 b,线性不等式

$$Ax\le b \tag{9.4}$$

的所有基本解是整数向量,那么 A 必是全幺模矩阵.

定理证明从略.

出现在线性规划运输问题中有一类重要的全幺模矩阵,其矩阵 $B=[b_{ij}]\in\mathbb{R}^{m\times n}$ 的特征描述如下:

(1) $b_{ij}=0,1$ 或 -1, $1\le i\le m$, $1\le j\le n$;

(2) B 的每一列至多包含两个非零元素;

(3) 存在指标集 $R_1,R_2\subset\{1,\cdots,m\}$,

$$R_1\cup R_2=\{1,\cdots,m\},\qquad R_1\cap R_2=\emptyset,$$

使得:如果在第 j 列中 $b_{ij}\ne 0$, $b_{kj}\ne 0$,那么

$$\begin{aligned}\operatorname{sgn} b_{ij}=\operatorname{sgn} b_{kj}&\Rightarrow \text{或者 } i\in R_1,k\in R_2,\\ &\qquad\text{或者 } k\in R_1,i\in R_2,\\ \operatorname{sgn} b_{ij}\ne\operatorname{sgn} b_{kj}&\Rightarrow \text{或者 } i,k\in R_1,\\ &\qquad\text{或者 } i,k\in R_2.\end{aligned}$$

换句话说,对应于 R_1 和 R_2, B 的行可以划分成不相交的两集.

例 矩阵

$$B=\begin{bmatrix}1&0&0&1&0&0\\0&1&0&0&1&0\\0&0&1&0&0&1\\1&1&1&0&0&0\\0&0&0&1&1&1\end{bmatrix}$$

是全幺模矩阵(证明留作练习).显然 B 具有特征(1)和(2).将 B 的行划分成前三行为一集,后两行为一集,亦即

$$R_1=\{1,2,3\},\qquad R_2=\{4,5\},$$

容易证明 B 满足特征(3).

10.9.8 定理 设 $A=[a_{ij}]\in\mathbb{R}^{m\times n}$ 是全幺模矩阵,则对 A 执行如下任一操作将不改变全幺模性:

(1) 行置换或列置换;

(2) 转置;

(3) 任一行或列乘以 -1;

(4) 增添只有一个元素或 1 或 -1 的行和列;

(5) 选主元 $a_{ij}\equiv\varepsilon\in\{1,-1\}$,对 A 交换第 i 行与第1行,第 j 列与第1列,所得矩阵记作

$$\begin{bmatrix}\varepsilon & r\\ c & \tilde{A}\end{bmatrix},$$

其中 $\tilde{A}\in\mathbb{R}^{(m-1)\times(n-1)}$.然后应用变换

$$\begin{bmatrix}\varepsilon & r\\ c & \tilde{A}\end{bmatrix}\mapsto\begin{bmatrix}-\varepsilon & \varepsilon r\\ \varepsilon c & \tilde{A}-\varepsilon cr\end{bmatrix}.$$

证明留作练习.

10.9.9 定义 实数域$\mathbb{R}$上线性空间X的子集L称为**格**,如果L满足下列性质:

(1) L对加法和减法封闭:

$$x, y \in L \Rightarrow x+y, x-y \in L.$$

(2) L是离散的:(按任一范数)X的任何有界集仅包含L的有限个点(向量).

例 格的一个基本例子是$\mathbb{R}^n$中分量$x_1, \cdots, x_n$为整数的点(向量)

$$x=\left(x_1, \cdots, x_n\right)^{\mathrm{T}}$$

的全体.

10.9.10 定义 设L是格.

如果L中存在线性无关向量组

$$x^{(1)}, \cdots, x^{(m)}, \tag{9.5}$$

使得L中任一向量都可表示成它们的具有整数系数的线性组合,则称(9.5)是L的一组**整数基**.

由L生成的线性空间的维数称为L的**维数**.

对于**10.9.9**的例中的格来说,$\mathbb{R}^n$中的标准单位基

$$e_1=(1,0, \cdots, 0)^{\mathrm{T}}, \cdots, e_n=(0, \cdots, 0,1)^{\mathrm{T}}$$

就是一组整数基,而且其维数为n.这个例子对下面的基本定理而言是很典型的.

10.9.11 定理 每个格有整数基.

证 设L是m维格,并设$x^{(1)}, \cdots, x^{(m)} \in L$是$L$生成线性空间的一组基.这样,每个$x \in L$可以唯一地表示为

$$x=\sum_{i=1}^m x_i x^{(i)}, \quad x_1, \cdots, x_m \in \mathbb{R}. \tag{9.6}$$

考虑L的子集

$$\tilde{L}=\left\{x=\sum_{i=1}^{m}x_i x^{(i)}\in L:0\le x_i\le 1,i=1,\cdots,m\right\},\quad (9.7)$$

$\tilde{L}$ 非空,因为(9.6)中 $x_1,\cdots,x_m$ 取0或1的每个向量都在 $\tilde{L}$ 中.由于 L 是离散的,$\tilde{L}$ 中只包含了 L 中的有限个向量.取 $y^{(1)}\in\tilde{L}$,

$$y^{(1)}=\sum_{i=1}^{m}\hat{x}_i x^{(i)},\quad (9.8)$$

使得

$$\hat{x}_1=\min\left\{x_1>0:x=\sum_{i=1}^{m}x_i x^{(i)}\in\tilde{L}\right\}.\quad (9.9)$$

现在,在基 $x^{(1)},\cdots,x^{(m)}$ 中用 $y^{(1)}$ 代替 $x^{(1)}$;于是每个 $x\in L$ 可以唯一地表示为

$$x=y_1 y^{(1)}+\sum_{i=2}^{m}y_i x^{(i)},\ \ y_1,\cdots,y_m\in\mathbb{R},\quad (9.10)$$

如此 y_1 必为整数.事实上,如果 y_1 为非整数,那么可以选取整数 l,使得

$$0<y_1-l<1.$$

显然,

$$x-ly^{(1)}=(y_1-l)y^{(1)}+\sum_{i=2}^{m}y_i x^{(i)}\in L.\quad (9.11)$$

用(9.8)代换等式(9.11)右边的 $y^{(1)}$,有

$$x-ly^{(1)}=(y_1-l)\hat{x}_1 x^{(1)}+\sum_{i=2}^{m}\left((y_1-l)\hat{x}_i+y_i\right)x^{(i)},\quad (9.12)$$

其中 $0<(y_1-l)\hat{x}_1<1$.再适当选取整数 $l_2,\cdots,l_m$,使得

$$0\le l_i+(y_1-l)\hat{x}_i+y_i\le 1,\quad i=2,\cdots,m.$$

因此,从(9.12)有

$$x - ly^{(1)} + \sum_{i=2}^{m} l_i x^{(i)}$$

$$= (y_1 - l)\hat{x}_1 x^{(1)} + \sum_{i=2}^{m} (l_i + (y_1 - l)\hat{x}_i + y_i) x^{(i)} \in \tilde{L}.$$

但是$(y_1 - l)\hat{x}_1 < \hat{x}_1$,这与$\hat{x}_1$的选取(9.9)矛盾.

至此,对格的维数m应用归纳法,便可完成定理的证明.用L_0表示形如(9.10)但$y_1 = 0$的向量x组成的集合.显然L_0是L的$m-1$维子格;由归纳法假设,L_0有整数基$y^{(2)},\cdots,y^{(m)}$.然后,由(9.10),$y^{(1)},\cdots,y^{(m)}$是L的整数基. □

10.9.12 定理 设L是$\mathbb{R}^n$中的n维格;列向量

$$x^{(1)},\cdots,x^{(n)} \quad 和 \quad y^{(1)},\cdots,y^{(n)}$$

是L的两组整数基;并设矩阵

$$X = \left[x^{(1)} \quad \cdots \quad x^{(n)}\right] \quad 和 \quad Y = \left[y^{(1)} \quad \cdots \quad y^{(n)}\right],$$

则存在幺模矩阵M,使得

$$X = MY.$$

证 依整数基的定义 **10.9.10**,每个向量$x^{(i)}$可以表示成$y^{(1)},\cdots,y^{(n)}$的具有整数系数的线性组合,显然存在具有整数元素的矩阵M,使得$X = MY$.因而从

$$\det X = \det M \det Y \neq 0,$$

推出$\det M \neq 0$即M可逆,有

$$Y = M^{-1}X.$$

另一方面,每个向量$y^{(i)}$可以表示成$x^{(1)},\cdots,x^{(n)}$的具有整数系数的线性组合;由此可知,M^{-1}也是具有整数元素的矩阵.这样,$\det M$和$\det M^{-1}$均为整数,而且

$$\det M^{-1} = \frac{1}{\det M},$$

于是必有$\det M = \pm 1$.从而M是幺模矩阵. □

这个定理表明整数基远非唯一.

整数矩阵的理论和幺模矩阵的研究成果可以参看[N1972],[B1971],[B1984]和[S1986].

10.10 纠错码组和奇偶校验矩阵

10.10.1 引言 在电子传送信息时经常发生干扰,收到的信息有可能是不正确的.编码理论的目标在于发现,而且如果可能的话,还有纠正发生在传送中的错误.常见的纠错码组的范例是 International Standard Book Number(ISBN),这是一种非常简单的算术校验,可用来确定一个给定的十进制数是否确实是一个 ISBN.这里仅考虑二**进制代码**,亦即由**代码字**(codeword)组成的集合,每个字是0与1构成的 n 位的**行**(string).

例 (1) 假定传送信息"北","南","东","西",其编码如下:

	北	南	东	西
代码 C_1	00	01	10	11
代码 C_2	000	101	011	110

代码 C_1 只要求传送两位,但这种代码不能发现任何错误,如果有一位或两位出现错误,那么收到的是一个不正确的信息.代码 C_2 则可发现任何**单**位(三位中只有一位)错误.

比如,传送101有一位出错,收到的是001,或111,或100,三者都不是代码字,接收者便知传送出错.

如此代码尽管不能纠正错误,但揭示了如何以附加位形式将过剩信息添加于原始信息,能正确地发现传送中的错误,这正是代码理论的基本途径和目标.

通常,假定任何单位出错的概率非常小.从而,如果接收到不正确信息,那么几乎只含一位而非多位出错.实际中,发现出错未加纠正的代

码的用处仅在于接收者得到已经发现错误的信息的复制而已.

(2) 一种简单实用的**检错码**(error-detecting code)是**奇偶校验**(parity check),由附加单位构成,使得整体代码字具有**偶数同位**(even parity),即具有偶数个1.接收到的字如果具有偶数同位,那么最有可能的情况是没有错误;如果具有奇数同位,则推断它有一位出错.

比如,发送的信息是101011,因偶数个1,故附加的校验位是0,给定的代码字是1010110;如果接收到的是1011110,那么因其含奇数个1,可断定在传送中(至少)一位出错.

但是,使用这种代码没有确定校正信息的途径.

(3) 假定传送的代码字是 $x_1 \cdots x_n$,其中每位是0或1,而且当

$$s \equiv x_1 + \cdots + x_{n-1}$$

是奇数时 $x_n = 1$,s 是偶数时 $x_n = 0$.如果接收的字是 $\hat{x}_1 \cdots \hat{x}_n$ 且有错误,那么依照 $\hat{x}_n$ 的奇偶性相同或不同于 $\hat{x}_1 \cdots \hat{x}_{n-1}$ 的奇偶性表明其分别存在偶数个位或奇数个位出错.然而,在多于一位出错的概率极小的假设下,本例中最有可能的情况分别是没有出错或一位出错.

(4) 传送信息由单位0或1 ('Yes'或'No')组成.长为3的**重复码**(repetition code)简单地由传送信息三次组成:

信　息:	0	1
代码字:	000	111

比如,接收到100,仍假定传送中至多一位出错,推断发送的是000.

通过全面考察收到信息的其余7种可能情形,即

000, 001, 010, 011, 101, 110, 111

表明如此代码全是校正单位错误.

10.10.2 定义　模2算术运算的**加法**和**乘法**如下:

$$0+0=0,\quad 0+1=1+0=1,\quad 1+1=0;$$
$$0\times0=0,\quad 0\times1=1\times0=0,\quad 1\times1=1. \tag{10.1}$$

为了赋予这些规则以意义,想象“1”表示奇数,“0”表示偶数,从而比如$1+1=0$意味着两个奇数之和是偶数.

还可定义**减法**为:

$$0-0=0,\quad 1-1=0,\quad 1-0=1,\quad 0-1=(1+1)-1=1.$$

两个代码字$a=a_1\cdots a_n$与$b=b_1\cdots b_n$之**和**是代码字,记作

$$a+b.$$

其第i位为

$$a_i+b_i,\quad 1\le i\le n,$$

其中加法遵循(10.1)的规则.

代码C称为**线性码**(linear code),如果C中任何两个代码字之和均属于C,即

$$a+b\in C,\quad \forall a,b\in C. \tag{10.2}$$

容易证明(留作练习):

(1) **10.10.1** 的**例**(1)中的代码

$$C=\{00,01,10,11\}$$

是线性码.

(2) 设a是线性码中的任一代码字,则$a+a=0$(每位均为零的代码字).因此0是任何线性码中的一个代码字.

10.10.3 定义 设$x=x_1\cdots x_n$是长为n的代码字,记

$$x'=(x_1,\cdots,x_n)^{\mathrm{T}}. \tag{10.3}$$

设H是$m\times n$的**二元矩阵**(binary matrix), $n>m$, 其所有元素是0或1.

由 H 可以**确定一个线性码**

$$C \equiv \left\{ x = x_1 \cdots x_n : Hx' = 0, x' = \left(x_1, \cdots, x_n\right)^{\mathrm{T}} \in \{0,1\}^n \right\}, \tag{10.4}$$

其中矩阵-向量乘法遵循(10.1)的规则.事实上,若

$$Hx' = 0, \qquad Hy' = 0,$$

则显然有

$$H(x' + y') = 0.$$

这表明 C 确是线性码.

H 称为代码 C 的**奇偶校验矩阵**(parity check matrix)或简称**校验矩阵**.

考虑校验矩阵形如

$$H = \begin{bmatrix} I_m & A \end{bmatrix}, \tag{10.5}$$

其中 $A = [a_{ij}]$ 是任意的 $m \times (n-m)$ 二元矩阵.此时,注意到减法规则 $0 - 1 = 1$,方程

$$Hx' = 0,$$

可以写成

$$\begin{cases} x_1 = a_{11} x_{m+1} + \cdots + a_{1,n-m} x_n \\ \cdots \\ x_m = a_{m1} x_{m+1} + \cdots + a_{m,n-m} x_n \end{cases}$$

或者

$$\begin{bmatrix} x_1 \\ \vdots \\ x_m \end{bmatrix} = A \begin{bmatrix} x_{m+1} \\ \vdots \\ x_n \end{bmatrix}. \tag{10.6}$$

因此,可以根据 $x_{m+1}, \cdots, x_n$ 各位独立取定的值,确定 $x_1, \cdots, x_m$ 各位的值.

$x_{m+1}, \cdots, x_n$ 各位称为**信息位**(information bit).

$x_1, \cdots, x_m$ 各位称为**校验位**(chech bit).

由于信息位每位可取0或1,有2^{n-m}种选取,故共有2^{n-m}个代码字.

信息位的位数$n-m$称为代码C的**维数**.

一般,对于任意的$m\times n$校验矩阵H来说,代码的维数等于$n-\operatorname{rank} H$.

注意,校验矩阵H中列的顺序是不重要的,H的列的重新排序只导致每个代码字中各位的重新排序.

10.10.4 定理 设$H\in\{0,1\}^{m\times n}$是代码C的校验矩阵,而且它既没有元素全为零的列,也没有恒同的两列,则C是**单位纠错码**(single-erroc-correcting code),亦即在每个代码字至多有一位出错的假设下,能发现从而纠正所有错误.

证 如果代码字$c\in C$在传送中出现单位错误,那么接收到的字形如

$$r=c+e_i,$$

其中e_i是第i位为1其余各位为0的字.

现在利用(10.3)中的记号.注意到$Hc'=0$,得

$$Hr'=Hc'+He_i'=He_i',$$

而He_i'正好是H的第i列.

由此,产生将接收信息r**译码**(recoding)成正确的传送代码字c的下述方案:

首先产生所谓**伴随式**(syndrome)Hr',无非有为零和不为零两种情况.

如果Hr'为零,那么r是代码字.

否则,即Hr'不为零,将它与H的列比较.若Hr'等于H的第i列,则出现单位传送错误,而且是其第i位出错;若Hr'不等于H的每一列,则发现了必在传送中多于一位出错. □

例 考虑校验矩阵

$$H=\begin{bmatrix}1&0&0&1&1\\0&1&0&0&1\\0&0&1&1&0\end{bmatrix}\equiv[I_3\quad A],\qquad(10.7)$$

其中

$$A\equiv\begin{bmatrix}1&1\\0&1\\1&0\end{bmatrix}.$$

由 H 定义长为5的线性码.设

$$x=x_1x_2x_3x_4x_5$$

是代码字,则 $Hx'=0$,具体地,

$$\begin{cases}x_1+\qquad\qquad x_4+x_5=0,\\ \qquad x_2+\qquad\qquad x_5=0,\\ \qquad\qquad x_3+x_4\qquad=0.\end{cases}$$

由此并注意到 x_1,x_2,x_3,x_4,x_5 是二进制数位,根据(10.1),有

$$\begin{cases}x_1=-x_4-x_5=x_4+x_5,\\ x_2=-x_5=x_5,\\ x_3=-x_4=x_4.\end{cases}$$

这里,x_4,x_5 是信息位,x_1,x_2,x_3 是校验位.于是

信息位(给定)		校验位(解得)			代码字				
x_4	x_5	x_1	x_2	x_3	x_1	x_2	x_3	x_4	x_5
0	0	0	0	0	0	0	0	0	0
0	1	1	1	0	1	1	0	0	1
1	0	1	0	1	1	0	1	1	0
1	1	0	1	1	0	1	1	1	1

现在假设接收字 $r=01101$，其伴随式

$$Hr'=H\begin{bmatrix}0\\1\\1\\0\\1\end{bmatrix}=\begin{bmatrix}1\\0\\1\end{bmatrix},$$

恒同 H 的第 4 列，这表明 r 的第 4 位单位出错.纠正译码信息是 01111.

另一情况，设 $r=00011$，则

$$Hr'=\begin{bmatrix}0\\1\\1\end{bmatrix},$$

不同于 H 的每一列，表明 r 必多位出错.

10.10.5 定义 设 $H\in\{0,1\}^{m\times n}$ 是校验矩阵，

$$m>1,\ m<n\le 2^m-1,$$

其列依此是 $1,2,\cdots,n$ 的二进制数表示，则称 H 是 **Hamming 校验矩阵**，称由 H 确定的线性码为 **Hamming 码**.

依 **10.10.4**，Hamming 码是单位纠错码.

当 $n=2^m-1$ 时，Hamming 码称为是**完全的(perfect)**.原因是其每个非零伴随式代表一个单位可纠正错误.

当 $n<2^m-1$ 时，Hamming 码称为是**不完全**的.原因是其某些非零伴随式代表多位错误，但却未能判定出错的位，自然说不上纠正错误.

例 以 $1,\cdots,6$ 的二进制数表示为列的 3×6 Hamming 校验矩阵

$$H=\begin{bmatrix}0&0&0&1&1&1\\0&1&1&0&0&1\\1&0&1&0&1&0\end{bmatrix}$$

确定的是不完全 Hamming 码,因为

$$n=6<2^m-1=2^3-1=7.$$

若接收字 $r=000111$,则伴随式是

$$Hr'=\begin{bmatrix}1\\1\\1\end{bmatrix},$$

而111是7的二进制数表示;由于字长为6,H 不包含以7的二进制数表示的列,故意味着发生多位传送错误.

如果将7的二进制数表示增添在 H 的各列之后,那么所得 3×7 Hamming 校验矩阵

$$\begin{bmatrix}0&0&0&1&1&1&1\\0&1&1&0&0&1&1\\1&0&1&0&1&0&1\end{bmatrix}$$

确定的便是完全 Hamming 码.

10.10.6 定义 设 $H\in\{0,1\}^{m\times n}$ 是 Hamming 校验矩阵.容易看出, H 经过列交换可以变换为形如(10.5)的标准形.现在,假设 H 已经是如此标准形,即

$$H=\begin{bmatrix}I_m & A\end{bmatrix},$$

这样,因 $Hx'=0$ 等价于(10.6),进而有扩展形式

$$\begin{bmatrix}x_1\\\vdots\\x_n\end{bmatrix}=\begin{bmatrix}A\\I_{n-m}\end{bmatrix}\begin{bmatrix}x_{m+1}\\\vdots\\x_n\end{bmatrix}.$$

此方程的转置形式是

$$(x_1,\cdots,x_n)=(x_{m+1},\cdots,x_n)\begin{bmatrix}A^{\mathrm{T}} & I_{n-m}\end{bmatrix}. \tag{10.8}$$

这是任一代码字关于信息位的直接表示,表明代码字全体是矩阵

$$G\equiv\begin{bmatrix}A^{\mathrm{T}} & I_{n-m}\end{bmatrix} \tag{10.9}$$

各行的所有可能的二元组合(即以0或1为系数的线性组合).

G 称为代码的**生成矩阵**(generated matrix).也可以换一种说法: G 的行构成代码字集合的一组**基**.

例 对于(10.7)中的校验矩阵 H ,代码生成矩阵是

$$G=\begin{bmatrix}A^{\mathrm{T}} & I_{n-m}\end{bmatrix}=\begin{bmatrix}1 & 0 & 1 & 1 & 0\\ 1 & 1 & 0 & 0 & 1\end{bmatrix},$$

其代码全体为

$$\alpha\times 10110+\beta\times 11001,\quad \forall\alpha,\beta\in\{0,1\}.$$

关于代码理论的进一步内容,可以参看[H1986]和[MS1977].

10.11 几种范数和几乎正规矩阵

10.11.1 定义 $\mathbb{C}^n$ 上的范数 $\|\bullet\|$ 称为**酉不变**(unitarily invariant)**向量范数**,如果对于所有的 $x\in\mathbb{C}^n$ 和所有的酉矩阵 $U\in\mathbb{C}^{n\times n}$,成立

$$\|Ux\|=\|x\|. \tag{11.1}$$

10.11.2 定理 在$\mathbb{C}^n$ 中,

(1) 2-范数(Euclid 范数) $\|\bullet\|_2$ 是酉不变的;

(2) 若 $\|\bullet\|$ 是酉不变向量范数,则存在某常数 $\alpha>0$,使得

$$\|\bullet\|=\alpha\|\bullet\|_2 .$$

(3) $\|\bullet\|_2$ 是满足 $\|e_1\|=1$ 的唯一酉不变向量范数.

证 (1) 对于任一 $x\in\mathbb{C}^n$ 和任一酉矩阵 $U\in\mathbb{C}^{n\times n}$,

$$\begin{aligned}\|Ux\|_2^2 &= (Ux,Ux) = (Ux)^{\mathrm{T}}\overline{(Ux)} = x^{\mathrm{T}}U^{\mathrm{T}}\overline{U}\overline{x} \\ &= x^{\mathrm{T}}\left(U^*U\right)^{\mathrm{T}}\overline{x} = x^{\mathrm{T}}\overline{x} = \|x\|_2^2.\end{aligned}$$

推出 $\|Ux\|_2 = \|x\|_2$.

(2)和(3)的证明从略. □

10.11.3 定义 $\mathbb{C}^{m\times n}$ 上的范数 $\|\bullet\|$ 称为**酉不变矩阵范数**,如果它对于所有的 $A\in\mathbb{C}^{m\times n}$ 和所有的酉矩阵 $U\in\mathbb{C}^{m\times m}$,酉矩阵 $V\in\mathbb{C}^{n\times n}$,成立

$$\|UAV\| = \|A\|. \tag{11.2}$$

熟知的酉不变矩阵范数的例子是 Frobenius 范数和谱范数.

10.11.4 定理 在 $\mathbb{C}^{m\times n}$ 中,

(1) Frobenius 范数 $\|\bullet\|_F$ 是酉不变的;而且

若 $A=[a_{ij}]\in\mathbb{C}^{m\times n}$,

$$\sigma_1 \geq \cdots \geq \sigma_q \geq 0 \ (q=\min\{m,n\})$$

是 A 的奇异值,则

$$\|A\|_F = \left(\sum_{i=1}^{m}\sum_{j=1}^{n}|a_{ij}|^2\right)^{1/2} = \left(\sum_{i=1}^{q}\sigma_i^2\right)^{1/2}. \tag{11.3}$$

(2) 谱范数 $\|\bullet\|_2$ 是酉不变的.

证 将 $A=[a_{ij}]\in\mathbb{C}^{m\times n}$ 写成

$$A=\begin{bmatrix}a^{(1)} & \cdots & a^{(n)}\end{bmatrix},$$

其中 $a^{(j)}\in\mathbb{C}^m$ 是 A 的第 j 列, $j=1,\cdots,n$.依 **1.8.29**,

$$\|A\|_F^2=\|a^{(1)}\|_2^2+\cdots+\|a^{(n)}\|_2^2.$$

又依 **10.11.2**,$\mathbb{C}^m$ 中的 2 -范数是酉不变的;于是,对于任何酉矩阵 $U\in\mathbb{C}^{m\times m}$,成立

$$\begin{aligned}\|UA\|_F^2&=\|Ua^{(1)}\|_2^2+\cdots+\|Ua^{(n)}\|_2^2\\&=\|a^{(1)}\|_2^2+\cdots+\|a^{(n)}\|_2^2=\|A\|_F^2.\end{aligned}$$

由此,并注意到对所有 $B\in\mathbb{C}^{m\times n}$ 有

$$\|B^*\|_F=\|B\|_F.$$

推出对于任何酉矩阵 $U\in\mathbb{C}^{m\times m}$ 和酉矩阵 $V\in\mathbb{C}^{n\times n}$,成立

$$\|UAV\|_F=\|AV\|_F=\|V^*A^*\|_F=\|A^*\|_F=\|A\|_F. \quad (11.4)$$

现在假设 $\sigma_1\geq\cdots\geq\sigma_q\geq 0\ (q=\min\{m,n\})$ 是 A 的奇异值.依 **5.4.5**,存在酉矩阵 $U\in\mathbb{C}^{m\times m}$ 和酉矩阵 $V\in\mathbb{C}^{n\times n}$,使得

$$A=U\Sigma V^*,$$

其中

$$\Sigma=\begin{bmatrix}\Sigma_q\\0\end{bmatrix}\text{ 或 }\begin{bmatrix}\Sigma_q & 0\end{bmatrix},\qquad \Sigma_q\equiv\operatorname{diag}(\sigma_1,\cdots,\sigma_q).$$

利用(11.4),得

$$\|A\|_F=\|U\Sigma V^*\|_F=\|\Sigma\|_F=\left(\sum_{i=1}^{n}\sigma_1^2\right)^{1/2}.$$

至此,(1)得证.

剩下证明(2).依谱范数定义和 $\mathbb{C}^n$ 中 2 -范数 $\|\cdot\|_2$ 是酉不变的,对于任何酉矩阵 $U\in\mathbb{C}^{m\times m}$ 和酉矩阵 $V\in\mathbb{C}^{n\times n}$,有

$$\|UAV\|_2=\max_{y\neq0}\frac{\|UAVy\|_2}{\|y\|_2}=\max_{y\neq0}\frac{\|AVy\|_2}{\|y\|_2}$$

$$= \max_{y \neq 0} \frac{\|A(Vy)\|_2}{\|Vy\|_2} = \max_{y \neq 0} \frac{\|Az\|_2}{\|z\|_2} = \|A\|_2,$$

其中 $y, z \in \mathbb{C}^n$. □

10.11.5 定义 设$\|\bullet\|$是$\mathbb{C}^n$上的向量范数.如果成立

$$|x| \le |y| \Rightarrow \|x\| \le \|y\|, \quad \forall x, y \in \mathbb{C}^n, \tag{11.5}$$

则称$\|\bullet\|$为**单调范数**(monotone norm).如果成立

$$\|x\| = \|\,|x|\,\|, \quad \forall x \in \mathbb{C}^n, \tag{11.6}$$

则称$\|\bullet\|$为**绝对范数**(absolute norm).

显然,所有 p -范数$\|\bullet\|_p$仅依赖 x 的分量的绝对值,而且是分量绝对值的增函数.因此所有 p -范数是单调范数和绝对范数.

10.11.6 定理 设$\|\bullet\|$是$\mathbb{C}^n$上的向量范数,$\|\bullet\|$为单调范数的充分必要条件是$\|\bullet\|$为绝对范数.

证 若$\|\bullet\|$为单调范数,$x \in \mathbb{C}^n$,令

$$y \equiv |x|,$$

则有$|y| \le |x|$,而且$|x| \le |y|$,推出$\|y\| \le \|x\|$,而且$\|x\| \le \|y\|$,因此

$$\|x\| = \|y\| \equiv \|\,|x|\,\|,$$

即$\|\bullet\|$为绝对范数.

反之,假设$\|\bullet\|$为绝对范数,$x = (x_1, \cdots, x_n)^{\mathrm{T}} \in \mathbb{C}^n$,$k$ 是整数,$1 \le k \le n$,$\alpha \in [0,1]$,则有

$$\begin{aligned}\|x\| &= \left\|(x_1, \cdots, x_{k-1}, \alpha x_k, x_{k+1}, \cdots, x_n)^{\mathrm{T}}\right\| \\ &= \left\|\frac{1}{2}(1-\alpha)(x_1, \cdots, x_{k-1}, -x_k, x_{k+1}, \cdots, x_n)^{\mathrm{T}} + \frac{1}{2}(1-\alpha)x + \alpha x\right\|\end{aligned}$$

$$\le \frac{1}{2}(1-\alpha)\left\|(x_1,\cdots,x_{k-1},-x_k,x_{k+1},\cdots,x_n)^{\mathrm{T}}\right\|$$

$$+\frac{1}{2}(1-\alpha)\|x\|+\alpha\|x\|$$

$$=\frac{1}{2}(1-\alpha)\|x\|+\frac{1}{2}(1-\alpha)\|x\|+\alpha\|x\|=\|x\|. \tag{11.7}$$

对不同分量重复(11.7),推出绝对范数$\|\cdot\|$满足

$$\left\|(\alpha_1 x_1,\cdots,\alpha_n x_n)^{\mathrm{T}}\right\|\le\left\|(x_1,\cdots,x_n)^{\mathrm{T}}\right\|,$$

$$\forall x=(x_1,\cdots,x_n)^{\mathrm{T}}\in\mathbb{C}^n,\ \ \alpha_1,\cdots,\alpha_n\in[0,1]. \tag{11.8}$$

由此,如果

$$|x|\le|y|,\quad x=(x_1,\cdots,x_n)^{\mathrm{T}},y=(y_1,\cdots,y_n)^{\mathrm{T}}\in\mathbb{C}^n,$$

那么必存在关系式

$$x_k=\alpha_k \mathrm{e}^{\mathrm{i}\theta_k}y_k,\quad \alpha_k\in[0,1],\theta_k\in\mathbb{R},\ k=1,\cdots,n.$$

于是,从(11.8)得

$$\|x\|=\left\|(\alpha_1\mathrm{e}^{\mathrm{i}\theta_1}y_1,\cdots,\alpha_n\mathrm{e}^{\mathrm{i}\theta_n}y_n)^{\mathrm{T}}\right\|$$

$$=\left\|(\alpha_1|y_1|,\cdots,\alpha_n|y_n|)^{\mathrm{T}}\right\|\le\left\|(|y_1|,\cdots,|y_n|)^{\mathrm{T}}\right\|=\|y\|,$$

即$\|\cdot\|$必为单调范数. □

从不等式(11.7)引出如下稍弱的单调性概念.

10.11.7 定义 $\mathbb{C}^n$上的范数$\|\cdot\|$称为**弱单调范数**(weakly monotone norm),如果

$$\left\|(x_1,\cdots,x_{k-1},0,x_{k+1},\cdots,x_n)^{\mathrm{T}}\right\|\le\|x\|,$$

$$\forall x=(x_1,\cdots,x_n)^{\mathrm{T}}\in\mathbb{C}^n,\quad k=1,\cdots,n. \tag{11.9}$$

如果向量范数$\|\cdot\|$是弱单调的,那么

$$\begin{aligned}
&\left\|\left(x_1,\cdots,x_{k-1},\alpha x_k,x_{k+1},\cdots,x_n\right)^{\mathrm{T}}\right\| \\
&=\left\|(1-\alpha)\left(x_1,\cdots,x_{k-1},0,x_{k+1},\cdots,x_n\right)^{\mathrm{T}}+\alpha x\right\| \\
&\le(1-\alpha)\left\|\left(x_1,\cdots,x_{k-1},0,x_{k+1},\cdots,x_n\right)^{\mathrm{T}}\right\|+\alpha\|x\| \\
&\le(1-\alpha)\|x\|+\alpha\|x\|=\|x\|, \\
&\qquad\qquad \forall x=\left(x_1,\cdots,x_n\right)^{\mathrm{T}}\in\mathbb{C}^n,\ \alpha\in[0,1].
\end{aligned}$$

这表明弱单调范数满足表面上较强的条件(11.8).

单调范数显然是弱单调范数,但反之不真.

例 设

$$f(x)\equiv\left|x_1-x_2\right|+\left|x_2\right|,\quad x=\left(x_1,x_2\right)^{\mathrm{T}}\in\mathbb{R}^2.$$

容易看出函数f满足正定性,次加性和齐次性,因此是$\mathbb{R}^2$上的范数.但是,当$x_1x_2<0$时,

$$f(|x|)=\left||x_1|-|x_2|\right|+\left||x_2|\right|\ne\left|x_1-x_2\right|+\left|x_2\right|=f(x).$$

由此可见,f不是绝对范数,亦即不是单调范数.而且,还容易看出f也不是弱单调范数.

弱单调范数但不是单调范数的例子可参看[HJ1985].

10.11.8 定义 $\mathbb{C}^{m\times n}$上的函数称为**广义矩阵范数**(generalized matrix norm),如果它满足矩阵范数定义 **1.8.28** 中的正定性、齐次性和次加性三个条件,而不包括次乘性.一般,仍采用记号$\|\cdot\|$.

自然,广义矩阵范数涵盖了矩阵范数.

10.11.9 定义 $\mathbb{C}^{m\times n}$上的广义矩阵范数$\|\cdot\|$称为 **Hadamard 次乘性范数**(submultiplicative norm),如果成立

$$\|A \circ B\| \le \|A\|\|B\|, \quad \forall A, B \in \mathbb{C}^{m\times n},$$

其中 $A \circ B$ 是 A 与 B 的 Hadamard 积.

10.11.10 定理 对于任何 $A, B \in \mathbb{C}^{m\times n}$,成立

$$\sigma_1(A \circ B) \le \sigma_1(A)\sigma_1(B),$$

这里 $\sigma_1(A \circ B), \sigma_1(A), \sigma_1(B)$ 分别是 $A \circ B, A, B$ 的最大奇异值.

定理证明从略.

这个定理说明谱范数是 Hadamard 次乘性范数.

10.11.11 定义 $\mathbb{C}^{n\times n}$ 上的广义矩阵范数 $\|\bullet\|$ 称为**谱优势范数**(spectrally dominant norm),如果

$$\|A\| \ge \rho(A), \quad \forall A \in \mathbb{C}^{n\times n},$$

这里 $\rho(A)$ 是 A 的谱半径.

$\mathbb{C}^{m\times n}$ 上的广义矩阵范数 $\|\bullet\|$ 称为**奇异值优势范数**(singular value dominant norm),如果

$$\|A\| \ge \sigma_1(A), \quad \forall A \in \mathbb{C}^{m\times n},$$

这里 $\sigma_1(A)$ 是 A 的最大奇异值.

如果 $\mathbb{C}^{n\times n}$ 上广义矩阵范数或者具有次乘性,或者是奇异值优势的,那么必是谱优势的.但是逆命题一般不成立.

10.11.12 定理 设 $\|\bullet\|$ 是 $\mathbb{C}^{n\times n}$ 上酉不变范数,则下列条件等价:

(1) $\|\bullet\|$ 是奇异值优势范数.

(2) $\|\bullet\|$ 是 Hadamard 次乘性范数.

(3) $\|\bullet\|$ 是谱优势范数.

(4) $\|\bullet\|$ 是次乘性范数,即 $\|\bullet\|$ 是矩阵范数.

定理证明从略.

10.11.13 定义 设 $A \in \mathbb{C}^{n\times n}$, $\|\bullet\|$ 是 $\mathbb{C}^{n\times n}$ 上的酉不变范数.记

$$\delta(A;\|\bullet\|) \equiv \inf\{\|T\| : A = U(\Lambda+T)U^*, U \text{ 是酉矩阵},$$

$$\Lambda \text{ 是对角矩阵}, T \text{ 是严格上三角矩阵}\}. \qquad (11.10)$$

依 **5.3.3**, $\delta(A;\|\bullet\|)$ 是有意义的,称之为 A 关于 $\|\bullet\|$ 的**距正规性的亏量**(defect from normality).如果 $\delta(A;\|\bullet\|)$ 是小的,则称 A 为**几乎正规矩阵**(almost normal matrix).

采用(11.10)作为非正规性的度量是基于:当 A 是正规矩阵时,依 **2.4.4**, $\delta(A;\|\bullet\|)=0$.

Frobenius 范数 $\|\bullet\|_F$ 和谱范数 $\|\bullet\|_2$ 是酉不变范数.

下面考虑 $\delta(A;\|\bullet\|_F)$ 和 $\delta(A;\|\bullet\|_2)$.

10.11.14 定理 设 $A \in \mathbb{C}^{n\times n}$. A 的特征值为

$$\lambda_1,\cdots,\lambda_n, \qquad |\lambda_1| \geq \cdots \geq |\lambda_n|.$$

A 的奇异值为

$$\sigma_1 \geq \cdots \geq \sigma_n,$$

则成立

$$\delta(A;\|\bullet\|_F)^2 = \sum_{i=1}^{n}\left(\sigma_i^2 - |\lambda_i|^2\right) \qquad (11.11)$$

和

$$|\sigma_i - |\lambda_i|| \leq \delta(A;\|\bullet\|_2), \quad i=1,\cdots,n. \qquad (11.12)$$

证 依 **5.3.3**,存在酉矩阵 $U \in \mathbb{C}^{n\times n}$,使得

$$A = U(\Lambda+T)U^*, \qquad (11.13)$$

其中

$$\Lambda = \mathrm{diag}(\lambda_1,\cdots,\lambda_n),$$

而$T\in\mathbb{C}^{n\times n}$是严格上三角矩阵.

另一方面,依 **5.4.5**,存在酉矩阵$V,W\in\mathbb{C}^{n\times n}$,使得

$$A=V\Sigma W^*,\tag{11.14}$$

其中

$$\Sigma=\mathrm{diag}(\sigma_1,\cdots,\sigma_n).$$

于是,利用$\|\bullet\|_F$的不变性,有

$$\|\Lambda+T\|_F=\|U(\Lambda+T)U^*\|_F=\|A\|_F=\|V\Sigma W^*\|_F=\|\Sigma\|_F.$$

由此,并注意到

$$\|\Lambda+T\|_F^2=\|\Lambda\|_F^2+\|T\|_F^2$$

和

$$\|\Lambda\|_F^2=\sum_{i=1}^n|\lambda_i|^2,\quad \|\Sigma\|_F^2=\sum_{i=1}^n\sigma_i^2,$$

得

$$\|T\|_F^2=\|\Sigma\|_F^2-\|\Lambda\|_F^2=\sum_{i=1}^n\left(\sigma_i^2-|\lambda_i|^2\right),\tag{11.15}$$

这表明$\|T\|_F^2$之值与U的选取无关.从而根据(11.10),有

$$\|T\|_F=\delta\left(A;\|\bullet\|_F\right).$$

联合此式和(11.15),即得(11.11).

现在证明(11.12).利用(11.13)和 **5.4.3**,知$\Lambda+T$和A有相同奇异值

$$\sigma_1\geq\cdots\geq\sigma_n\geq 0,$$

Λ的奇异值为

$$|\lambda_1|\geq\cdots\geq|\lambda_n|.$$

令

$$\tilde{S}\equiv\mathrm{span}\{e_i,\cdots,e_n\},$$

于是,依 **5.4.6**,

$$\begin{aligned}\sigma_i &= \min_{\dim S=n-i+1}\max_{x\in S,\|x\|_2=1}\left\|(\Lambda+T)x\right\|_2\\ &\le \max_{x\in\tilde S,\|x\|_2=1}\left\|(\Lambda+T)x\right\|_2\\ &\le \max_{x\in\tilde S,\|x\|_2=1}\|\Lambda x\|_2+\max_{x\in\tilde S,\|x\|_2=1}\|Tx\|_2\\ &\le \max_{x\in\tilde S,\|x\|_2=1}\|\Lambda x\|_2+\max_{\|x\|_2=1}\|Tx\|_2=|\lambda_i|+\|T\|_2.\end{aligned}\tag{11.16}$$

类似地,有

$$\begin{aligned}|\lambda_i| &= \min_{\dim S=n-i+1}\max_{x\in S,\|x\|_2=1}\|\Lambda x\|_2\\ &\le \min_{\dim S=n-i+1}\max_{x\in S,\|x\|_2=1}\left\|(\Lambda+T)x\right\|_2\\ &+\min_{\dim S=n-i+1}\max_{x\in S,\|x\|_2=1}\left\|(-T)x\right\|_2\le\sigma_i+\|T\|_2.\end{aligned}$$

联合此不等式和(11.16),得

$$\left|\sigma_i-|\lambda_i|\right|\le\|T\|_2.$$

由此及(11.10)推出(11.12). □

如果 A 是正规矩阵,那么依 **5.4.4**,其奇异值必等于特征值的绝对值,即

$$\sigma_i=|\lambda_i|,\quad i=1,\cdots,n.$$

此时, A 是酉可对角化的,有

$$\delta(A;\|\bullet\|)=0.$$

如果 A 非正规矩阵,那么(11.11)表明:

$$\sigma_i\approx|\lambda_i|,\quad i=1,\cdots,n\tag{11.17}$$

的充分必要条件是 A 为在 $\delta(A;\|\bullet\|_F)$ 很小意义下的几乎正规矩阵.

而(11.12)是说:若 A 为在 $\delta(A;\|\bullet\|_2)$ 很小意义下的几乎正规矩阵,则条件(11.17)成立.

10.12 对合矩阵和共轭对合矩阵

10.12.1 定义 $A \in \mathbb{C}^{n\times n}$ 称为**对合矩阵**(involutory matrix),如果

$$(I-A)(I+A)=0. \tag{12.1}$$

10.12.2 定理 设 $A \in \mathbb{C}^{n\times n}$ 是对合矩阵,则

(1) 当 $n=2$ 时, A 形如

$$A=\begin{bmatrix} a & b \\ c & -a \end{bmatrix}, \quad a^2+bc=1$$

或者

$$A=I \quad 或 \quad A=-I.$$

(2) $B \equiv \frac{1}{2}(I+A)$ 是等幂矩阵.

证明留作练习.

10.12.3 定义 $E \in \mathbb{C}^{n\times n}$ 称为**共轭对合矩阵**(coninvolutory matrix)或**圆矩阵**(circular matrix),如果

$$E\overline{E}=I. \tag{12.2}$$

10.12.4 定理 $E \in \mathbb{C}^{n\times n}$ 为共轭对合矩阵的充分必要条件是存在 $B \in \mathbb{C}^{n\times n}$,使得

$$E=\overline{B}^{-1}B.$$

证 充分性.若 $E=\overline{B}^{-1}B$,则

$$E\overline{E}=\overline{B}^{-1}BB^{-1}\overline{B}=\overline{B}^{-1}\overline{B}=I.$$

必要性证明从略.可参看[HJ1991]的 478 页. □

10.12.5 定理　设 $E \in \mathbb{C}^{n\times n}$.下列条件等价:

(1) $E\overline{E} = I$,即 E 是共轭对合矩阵.

(2) $E = \mathrm{e}^{\mathrm{i}S}, S \in \mathbb{R}^{n\times n}$.

(3) $E = F^2, F \in \mathbb{C}^{n\times n}$ 是共轭对合矩阵.

证明从略.可参看[HJ1991]的 480 页.

10.12.6 定理　设 $A \in \mathbb{C}^{n\times n}$ 是非奇异矩阵,则

(1) 存在共轭对合矩阵 $E \in \mathbb{C}^{n\times n}$ 和非奇异矩阵 $R \in \mathbb{R}^{n\times n}$,使得

$$A = RE,$$

而且 E 是 $\overline{A}^{-1}A$ 的多项式.

(2) 如果 $A = RE$,其中满足 $E\overline{E} = I$, $R \in \mathbb{R}^{n\times n}$ 是非奇异矩阵,那么

$$E^2 = \overline{A}^{-1}A.$$

而且,若 R 和 E 可交换,则 A 和 $\overline{A}$ 可交换(即 $A\overline{A}$ 是实的);反之. 若 A 和 $\overline{A}$ 可交换, E 是 $\overline{A}^{-1}A$ 的多项式,则 R 和 E 可交换.

证　因 $\overline{A}^{-1}A$ 是共轭对合矩阵,**10.12.5** 保证它有共轭对合的平方根 E,亦即

$$\overline{A}^{-1}A = E^2, \quad E\overline{E} = I.$$

而且, E 可取为 $\overline{A}^{-1}A$ 的多项式(见 **9.3.7**).由 $\overline{A}^{-1}A = E^2$,有

$$A = \overline{A}E^2 = \left(\overline{A}E\right)E,$$

从而

$$A\overline{E} = \left(\overline{A}E\right)E\overline{E} = \overline{A}E = \overline{A\overline{E}}.$$

因此, $R \equiv A\overline{E}$ 是实的, $A = RE$.

现设 $A = RE$, R 是实的非奇异的, $E\overline{E} = I$.若 R 和 E 可交换,则有

$$A\overline{A} = RER\overline{E} = R\overline{E}RE = \overline{A}A = \overline{A\overline{A}},$$

故 A 和 $\overline{A}$ 可交换, $A\overline{A}$ 是实的.反之,若 A 和 $\overline{A}$ 可交换,则有

$$RER\overline{E} = A\overline{A} = \overline{A}A = R\overline{E}RE,$$

故 $ER\overline{E}=\overline{E}RE$,从而

$$E^2R=E\left(ER\overline{E}\right)E=E\left(\overline{E}RE\right)E=RE^2,$$

即 R 和 $E^2=\overline{A}^{-1}A$ 可交换.由此推出,如果 E 是 $\overline{A}^{-1}A$ 的多项式,那么 R 和 E 必可交换. □

10.12.7 定理 设 $A\in\mathbb{C}^{m\times n}$, $m\geq n$.存在 $R\in\mathbb{R}^{m\times n}$ 和共轭对合矩阵 $E\in\mathbb{C}^{n\times n}$ 使得

$$A=RE$$

的充分必要条件是非奇异矩阵 $S\in\mathbb{C}^{n\times n}$ 使得

$$A=\overline{A}S.$$

证 先进行分析.

如果 A 有所要求形式的分解 $A=RE$,而且若 $S\in\mathbb{C}^{n\times n}$ 非奇异,则

$$AS=RES=R\left(ES\right),$$

$ES\in\mathbb{C}^{n\times n}$ 是非奇异的.依 **10.12.6**,存在 $\tilde{R}\in\mathbb{R}^{n\times n}$ 和共轭对合矩阵 $\tilde{E}\in\mathbb{C}^{n\times n}$ 使得

$$ES=\tilde{R}\tilde{E}.$$

于是

$$AS=R\left(ES\right)=R\left(\tilde{R}\tilde{E}\right)=\left(R\tilde{R}\right)\tilde{E},\tag{12.3}$$

这表明关于 AS 也存在所要求形式的分解.

另一方面,若 $S,S_0\in\mathbb{C}^{n\times n}$ 非奇异, $A=\overline{A}S$,且 $A_0\equiv AS_0$,则

$$A_0=\overline{A}SS_0=\left(\overline{AS_0}\right)\left(\overline{S_0}^{-1}SS_0\right)=\overline{A_0}S_1,\tag{12.4}$$

其中 $S_1\equiv\overline{S_0}^{-1}SS_0$ 是非奇异的.

综合(12.3)和(12.4)得知,要推证的两个等价条件在右乘任何非奇异矩阵下都是不变的.

因为存在置换矩阵 $P\in\mathbb{R}^{n\times n}$,使得

$$AP\equiv\left[A_1\quad A_2\right],\quad \mathrm{rank}A=\mathrm{rank}A_1=r,\ \ A_1\in\mathbb{C}^{m\times r}.$$

又鉴于 P 非奇异并右乘于 A,根据上述讨论,可以直接假定

$$A \equiv \begin{bmatrix} A_1 & A_2 \end{bmatrix}, \quad \text{rank}A = \text{rank}A_1 = r, \quad A_1 \in \mathbb{C}^{m\times r}.$$

此时,A_2 的每一列可以表示成 A_1 各列的线性组合,从而必存在 $B \in \mathbb{C}^{r\times(n-r)}$,使得

$$A = \begin{bmatrix} A_1 & A_2 \end{bmatrix} = \begin{bmatrix} A_1 & A_1B \end{bmatrix} = \begin{bmatrix} A_1 & 0 \end{bmatrix}\begin{bmatrix} I & B \\ 0 & I \end{bmatrix}.$$

注意到其中

$$\begin{bmatrix} I & B \\ 0 & I \end{bmatrix} \in \mathbb{C}^{n\times n}$$

非奇异,并用其右乘$\begin{bmatrix} A_1 & 0 \end{bmatrix}$,同样根据上述讨论,下面可以进一步假设

$$A \equiv \begin{bmatrix} A_1 & 0 \end{bmatrix}, \quad \text{rank}A_1 = r, \quad A_1 \in \mathbb{C}^{m\times r}. \qquad (12.5)$$

现在,设对于某个非奇异矩阵 $S \in \mathbb{C}^{n\times n}$,使得 $A = \overline{A}S$.此恒等式在(12.5)假设下的分块形式如下:

$$\begin{bmatrix} A_1 & 0 \end{bmatrix} = A = \overline{A}S = \begin{bmatrix} \overline{A_1} & 0 \end{bmatrix}\begin{bmatrix} S_{11} & S_{12} \\ S_{21} & S_{22} \end{bmatrix} = \begin{bmatrix} \overline{A_1}S_{11} & \overline{A_1}S_{12} \end{bmatrix},$$

于是有

$$0 = \overline{A_1}S_{12}$$

和

$$A_1 = \overline{A_1}S_{11} = \overline{\left(\overline{A_1}S_{11}\right)}S_{11} = A_1\left(\overline{S_{11}}S_{11}\right)$$

或

$$A_1\left(I - \overline{S_{11}}S_{11}\right) = 0.$$

由此,因 $A_1 \in \mathbb{C}^{m\times r}$ 且 $\text{rank}A_1 = r$,推出

$$S_{12} = 0, \quad \overline{S_{11}}S_{11} = I,$$

表明 S_{11} 是共轭对合矩阵.依 **10.12.5** 的(3),$S_{11} = E^2$,E 是共轭对合矩阵.因此

$$A_1 = \overline{A_1}S_{11} = \overline{A_1}E^2 = RE,$$

其中

$$R = \overline{A_1}E = \overline{(\overline{A_1}E^2)}E = A_1\overline{E} = \overline{(\overline{A_1}E)}$$

是实矩阵.最后

$$A = [A_1 \quad 0] = [RE \quad 0] = [R \quad 0]\begin{bmatrix} E & 0 \\ 0 & I \end{bmatrix}.$$

这是所要求形式的分解.

反之,设 $A = RE$, $R \in \mathbb{R}^{m\times n}$,则因

$$\overline{A}E = \overline{(RE)}E = R\overline{E}E = R,$$

故有

$$A = RE = (\overline{A}E)E = \overline{A}E^2,$$

E^2 是非奇异的. □

10.13 自伴矩阵偏序及正定矩阵若干不等式

10.13.1 定义 设 $A, B \in \mathbb{C}^{n\times n}$ 是自伴矩阵.记号 $A \succcurlyeq B$ 是指 $A - B$ 是半正定的;记号 $A \succ B$ 是指 $A - B$ 是正定的.

可以认为自伴矩阵是实数的延伸,正定矩阵是正实数的延伸,自然会问在自伴矩阵之间有没有一种好的不等或(部分)序的概念.记号$\succcurlyeq$和$\succ$正是为此而引入的.

容易证明(留作练习):

(1) **10.13.1** 定义的关系"$\succcurlyeq$"和矩阵相等概念是相容的,也就是说,若 $A \succcurlyeq B$ 且 $B \succcurlyeq A$,则 $A = B$.

(2) 关系"$\succcurlyeq$"具有自反性

$$A \succcurlyeq A$$

和传递性

$$A \succcurlyeq B, B \succcurlyeq C \Rightarrow A \succcurlyeq C,$$

但“$\succcurlyeq$”不是全序,也就是说,存在大量的 $A, B \in \mathbb{C}^{n\times n}$,既非 $A \succcurlyeq B$ 也非 $B \succcurlyeq A$.

关系$\succcurlyeq$和$\succ$称为$\mathbb{C}^{n\times n}$中自伴矩阵的**偏序**.

10.13.2 定理 设 $A, B \in \mathbb{C}^{n\times n}$ 是自伴矩阵,则

(1) 如果 $A \succcurlyeq B$,那么

$$T^*AT \succcurlyeq T^*BT, \quad \forall T \in \mathbb{C}^{n\times m}. \tag{13.1}$$

(2) 如果 $A \succ B$, $m \le n$,那么

$$T^*AT \succ T^*BT, \quad \forall T \in \mathbb{C}^{n\times m}, \ \operatorname{rank} T = m. \tag{13.2}$$

证 若 $A - B$ 半正定,则

$$y^*(A-B)y \ge 0, \quad \forall y \in \mathbb{C}^n.$$

因此对任何 $T \in \mathbb{C}^{n\times m}$,

$$x^*(T^*AT - T^*BT)x = (Tx)^*(A-B)(Tx) \ge 0, \quad \forall x \in \mathbb{C}^m,$$

这表明 $T^*AT - T^*BT$ 是半正定的,即成立(13.1).(1)得证.

(2)的证明留作练习. □

10.13.3 定理 设 $A \in \mathbb{C}^{n\times n}$ 是正定矩阵, $B \in \mathbb{C}^{n\times n}$ 是半正定矩阵,则

(1) $A \succcurlyeq B$ 的充分必要条件是 $\rho(BA^{-1}) \le 1$.

(2) $A \succ B$ 的充分必要条件是 $\rho(BA^{-1}) < 1$.

证 依 **10.6.10** 和 **10.6.8**,存在非奇异矩阵 $C \in \mathbb{C}^{n\times n}$,使得

$$A = CIC^*, \qquad B = CDC^*, \ D \equiv \operatorname{diag}(d_1, \cdots, d_n).$$

由此,并注意到 C 的非奇异性,推出 $A \succcurlyeq B$ 当且仅当

$$C(I-D)C^* \ge 0.$$

显然,这后一条件又等价于

$$d_i \le 1, \quad i = 1, \cdots, n.$$

鉴于

$$BA^{-1}=CDC^*\left(C^*\right)^{-1}C^{-1}=CDC^{-1},$$

BA^{-1} 的特征值正好是 $d_1,\cdots,d_n$;而 $0\le d_i\le 1,i=1,\cdots,n$ 等价于 $\rho\left(BA^{-1}\right)\le 1$.(1)得证.

(2)的证明留作练习. □

10.13.4 推论 设 $A,B\in\mathbb{C}^{n\times n}$ 是正定矩阵,则

(1) $A\succcurlyeq B$ 的充分必要条件是 $B^{-1}\succcurlyeq A^{-1}$.

(2) 如果 $A\succcurlyeq B$,那么 $\det A\ge\det B$,且 $\mathrm{tr}A\ge\mathrm{tr}B$.

(3) 如果 $A\succcurlyeq B$, A 和 B 的特征值分别按相同(递增或递减)顺序排列,那么

$$\lambda_k(A)\ge\lambda_k(B),\quad k=1,\cdots,n.$$

证 依 **10.13.3**, $A\succcurlyeq B$ 当且仅当 $\rho\left(BA^{-1}\right)\le 1$.然而

$$\rho\left(BA^{-1}\right)=\rho\left(A^{-1}\left(BA^{-1}\right)A\right)=\rho\left(A^{-1}B\right).$$

仍依 **10.13.3**, $\rho\left(A^{-1}B\right)\le 1$ 当且仅当 $B^{-1}\succcurlyeq A^{-1}$.

若 $A\succcurlyeq B$,则 $\rho\left(BA^{-1}\right)\le 1$,且依 **10.6.8**, BA^{-1} 的所有特征值是正的,故它们必在区间 $(0,1]$ 内.于是,

$$\det\left(BA^{-1}\right)\le 1.$$

由此推出 $\det A\ge\det B$.另一方面,从 **10.13.3** 证明中得知

$$A=CC^*,\qquad B=CDC^*,$$

其中

$$C=[c_{ij}]\in\mathbb{C}^{n\times n},$$

而

$$D=\mathrm{diag}\left(d_1,\cdots,d_n\right),\quad 0<d_i\le 1,\ i=1,\cdots,n,$$

因此

$$\mathrm{tr}B=\mathrm{tr}CDC^*=\mathrm{tr}DC^*C=\sum_{i,j=1}^{n}d_i\left|c_{ij}\right|^2$$

$$\le \sum_{i,j=1}^{n} \left| c_{ij} \right|^2 = \mathrm{tr} CC^* = \mathrm{tr} A .$$

结论(3)(蕴涵(2)中的行列式不等式和迹不等式)直接从 **3.5.5** 推出. □

10.13.5 定理 设自伴矩阵 P 分块成

$$P = \begin{bmatrix} A & B \\ B^* & C \end{bmatrix}, \tag{13.3}$$

其中 A 和 C 是方阵,则 P 正定的充分必要条件是 A 正定且

$$C \succ B^* A^{-1} B .$$

而且,这后一条件等价于

$$\rho\left(B^* A^{-1} B C^{-1}\right) < 1 . \tag{13.4}$$

证 先证必要性. P 正定,则 A 和 C 是正定的,而且

$$P = \begin{bmatrix} \left(A - BC^{-1}B^*\right)^{-1} & A^{-1}B\left(B^* A^{-1} B - C\right)^{-1} \\ \left(B^* A^{-1} B - C\right)^{-1} B^* A^{-1} & \left(C - B^* A^{-1} B\right)^{-1} \end{bmatrix} \tag{13.5}$$

也是正定的.于是

$$\left(A - BC^{-1}B^*\right)^{-1}, \quad A - BC^{-1}B^*,$$
$$\left(C - B^* A^{-1} B\right)^{-1}, \quad C - B^* A^{-1} B,$$

都是正定的.从而有

$$A \succ 0,\ C \succ 0,\ A \succ BC^{-1}B^*,\ C \succ B^* A^{-1} B .$$

再依 **10.13.3** 的(2),即得(13.4).

再证充分性.设 A 正定而且 $C \succ B^* A^{-1} B$.令

$$X \equiv -A^{-1} B ,$$

则有

$$\begin{bmatrix} I & 0 \\ X^* & I \end{bmatrix} \begin{bmatrix} A & B \\ B^* & C \end{bmatrix} \begin{bmatrix} I & X \\ 0 & I \end{bmatrix} = \begin{bmatrix} A & 0 \\ 0 & C - B^* A^{-1} B \end{bmatrix}. \tag{13.6}$$

因为右端矩阵是正定的,而且*相合于矩阵 P ,依 **10.13.2** 的(2),推出

矩阵 P 是正定的. □

10.13.6 定理 设 $A \in \mathbb{C}^{n\times n}$ 和 $C \in \mathbb{C}^{m\times m}$ 是正定矩阵, $B \in \mathbb{C}^{n\times m}$, 则下列条件等价:

(1) $(x^* Ax)(y^* Cy) \geq |x^* By|^2, \forall x \in \mathbb{C}^n, y \in \mathbb{C}^m$.

(2) $x^* Ax + y^* Cy \geq 2|x^* By|, \forall x \in \mathbb{C}^n, y \in \mathbb{C}^m$.

(3) $\rho(B^* A^{-1} BC^{-1}) \leq 1$.

(4) $\begin{bmatrix} A & B \\ B^* & C \end{bmatrix} \succcurlyeq 0$.

证 (1)⇒(2):利用算术-几何平均不等式及(1),有

$$\frac{1}{2}(x^* Ax + y^* Cy) \geq (x^* Ax)^{1/2}(y^* Cy)^{1/2} \geq |x^* By|.$$

(2)⇒(3):(2)可以写成

$$x^* Ax + y^* Cy = (A^{1/2}x)^*(A^{1/2}x) + (C^{1/2}y)^*(C^{1/2}y) \geq 2|x^* By|,$$

由此,用 x 替换 $A^{1/2}x$, y 替换 $C^{1/2}y$,有

$$x^* x + y^* y \geq 2\left|(A^{-1/2}x)^* B(C^{-1/2}y)\right| = 2|x^* A^{-1/2} BC^{-1/2} y|.$$

在此不等式中,令

$$x \equiv A^{-1/2} BC^{-1/2} y,$$

得

$$y^* C^{-1/2} B^* A^{-1} BC^{-1/2} y + y^* y \geq 2|y^* C^{-1/2} B^* A^{-1} BC^{-1/2} y|. \tag{13.7}$$

因为 $C^{-1/2} B^* A^{-1} BC^{-1/2}$ 是半正定的,(13.7)等价于

$$y^* y \geq y^* C^{-1/2} B^* A^{-1} BC^{-1/2} y, \quad \forall y \in \mathbb{C}^m. \tag{13.8}$$

如果选取 y 是 $C^{-1/2} B^* A^{-1} BC^{-1/2}$ 的特征向量,那么不等式(13.8)是说对应的特征值(必为非负)以1为界.

于是有

$$
\begin{aligned}
1 &\geq \rho\left(C^{-1/2}B^{*}A^{-1}BC^{-1/2}\right) \\
&= \rho\left(C^{1/2}\left(C^{-1/2}B^{*}A^{-1}BC^{-1/2}\right)C^{-1/2}\right) \\
&= \rho\left(B^{*}A^{-1}BC^{-1}\right).
\end{aligned}
$$

(3)$\Rightarrow$(1):利用(3),对任何 $x \in \mathbb{C}^{n}$ 和 $y \in \mathbb{C}^{m}$,有

$$
\begin{aligned}
\left|x^{*}\left(A^{-1/2}BC^{-1/2}y\right)\right|^{2} &\leq \|x\|_{2}^{2}\left\|A^{-1/2}BC^{-1/2}y\right\|_{2}^{2} \\
&= \left(x^{*}x\right)\left(y^{*}C^{-1/2}B^{*}A^{-1}BC^{-1/2}y\right) \leq \left(x^{*}x\right)\left(y^{*}y\right),
\end{aligned}
$$

其中 $\|x\|_{2} \equiv \left(x^{*}x\right)^{1/2}$.在以上不等式中,用 x 替换 $A^{-1/2}x$,用 y 替换 $C^{-1/2}y$,得

$$
\left|x^{*}By\right|^{2} \leq \left(x^{*}Ax\right)\left(y^{*}Cy\right).
$$

(3)$\Leftrightarrow$(4):即 **10.13.5**. □

10.13.7 定理 设 $P \in \mathbb{C}^{n \times n}$ 是正定矩阵,$\alpha \subset \{1, \cdots, n\}$ 是指标集,则

$$
P^{-1}(\alpha) \succcurlyeq \left(P(\alpha)\right)^{-1}, \tag{13.9}
$$

其中左端是 P^{-1} 的主子矩阵,右端是 P 的主子矩阵之逆.

证 $\mathbb{C}^{n \times n}$ 中的正定矩阵集合在置换相合变换下是封闭的.不妨假设

$$
P = \begin{bmatrix} A & B \\ B^{*} & C \end{bmatrix}, \qquad P(\alpha) = A.
$$

因此,$\left(P(\alpha)\right)^{-1} = A^{-1}$,并由(13.5)得知

$$
P^{-1}(\alpha) = \left(A - BC^{-1}B^{*}\right)^{-1}.
$$

由于 $C \succ 0$,有 $BC^{-1}B^{*} \succcurlyeq 0$,再注意到 $A - BC^{-1}B^{*} \succ 0$,得

$$
A \succcurlyeq A - BC^{-1}B^{*} \succ 0.
$$

由此,从 **10.13.4** 的(1)推出(13.9). □

10.13.8 定理 设 $A,B\in\mathbb{C}^{n\times n}$ 是正定矩阵,则

(1) $A^{-1}\circ B^{-1}\succcurlyeq(A\circ B)^{-1}$.

(2) $A^{-1}\circ A^{-1}\succcurlyeq(A\circ A)^{-1}$.

(3) $A^{-1}\circ A\succcurlyeq I\succcurlyeq(A^{-1}\circ A)^{-1}$.

这里$\circ$是 Hadamard 乘法记号.

证 依 **7.2.3** 和 **7.1.6**,并将 **10.13.7** 应用于 $P\equiv A\otimes B$,有

$$A^{-1}\circ B^{-1}=\left(A^{-1}\otimes B^{-1}\right)(\alpha)=(A\otimes B)^{-1}(\alpha)$$

$$\geq\left[(A\otimes B)(\alpha)\right]^{-1}=(A\circ B)^{-1}.$$

这里指标集α的取法见 **7.2.3**.(1)得证.

(2)和(3)分别是(1)取 $B=A$ 和 $B=A^{-1}$ 的特殊情形.注意,在(3)中,$A^{-1}\circ A\succcurlyeq I$ 得自 **7.2.15**;然后由其及 **10.13.4** 的(1),即得 $I\succcurlyeq(A^{-1}\circ A)^{-1}$. □

10.13.9 Fischer 不等式 设

$$P=\begin{bmatrix}A & B\\ B^{*} & C\end{bmatrix}$$

是正定矩阵,A 和 C 是方阵,则

$$\det P\leq(\det A)(\det C).\tag{13.10}$$

证 利用(13.6),得

$$\det P=\det\begin{bmatrix}A & 0\\ 0 & C-B^{*}A^{-1}B\end{bmatrix}$$

$$=(\det A)\left(\det\left(C-B^{*}A^{-1}B\right)\right)\leq(\det A)(\det C).$$

最后的不等号利用了 **10.13.5** 和 **10.13.4** 的(2),因

$$C\succcurlyeq C-B^{*}A^{-1}B\succcurlyeq 0,$$

故必有 $\det C\geq\det\left(C-B^{*}A^{-1}B\right)$. □

10.13.10 Szasz 不等式 设 $A=[a_{ij}]\in\mathbb{C}^{n\times n}$ 是正定矩阵,$P_k(A)$表示

A 的所有 $k\times k$ 主子式——共有 $\binom{n}{k}$ 个——之积,注意

$$P_1(A)=a_{11}\cdots a_{nn},\quad P_n(A)=\det A,$$

则

$$P_{k+1}(A)^{\binom{n-1}{k}^{-1}}\le P_k(A)^{\binom{n-1}{k-1}^{-1}},\quad k=1,\cdots,n-1. \tag{13.11}$$

证 因为 A^{-1} 的对角元素正好是 A 的 $(n-1)\times(n-1)$ 主子式与 $\det A$ 之比,依 **3.10.3**,

$$\frac{1}{\det A}=\det A^{-1}\le\frac{P_{n-1}(A)}{(\det A)^n},$$

因此

$$P_n(A)^{n-1}=(\det A)^{n-1}\le P_{n-1}(A).$$

此不等式两边开 $n-1$ 次方,得(13.11)的 $k=n-1$ 的情形.其余的情形可以归纳地导出.例如,$k=n-2$ 的情形,将每个 $(n-1)\times(n-1)$ 主子矩阵视为原始矩阵,应用上述不等式,并注意到 A 的每个 $(n-2)\times(n-2)$ 主子矩阵是 A 的两个 $(n-1)\times(n-1)$ 主子矩阵的主子矩阵,得

$$P_{n-1}(A)^{n-2}\le P_{n-2}(A)^2,$$

然后对此不等式两边开 $(n-1)(n-2)$ 次方. □

10.13.11 引理 设 $A\in\mathbb{C}^{n\times n}$ 是半正定矩阵,定义

$$\alpha(A)=\begin{cases}\det A/\det A_{11}, & \text{当}A_{11}\text{是正定的},\\ 0, & \text{否则},\end{cases}$$

其中 A_{11} 是 A 删去第1行和第1列后的 $(n-1)\times(n-1)$ 主子矩阵.设 $E_{11}\in\mathbb{R}^{n\times n}$ 是 $(1,1)$-元素为1其余元素为0的矩阵,则矩阵

$$A-tE_{11}$$

对所有 $t\le\alpha(A)$ 是半正定的,对任何 $t>\alpha(A)$ 是非半正定的;特别,$A-\alpha(A)E_{11}$ 是半正定的.

证 注意到 $A-tE_{11}$ 与 A 所有不包含第1行和第1列的对应的主子式是相等的,而且

$$\det(A-tE_{11})=\det A-t\det A_{11},$$

并应用 **3.10.6**,即可推出当 A 正定时引理成立.由此再通过极限处理,得知当 A 半正定时结论也成立. □

10.13.12 Oppenheim 不等式 设 $A,B=[b_{ij}]\in\mathbb{C}^{n\times n}$ 是半正定矩阵,则

$$(\det A)\prod_{i=1}^{n}b_{ii}\le\det(A\circ B), \tag{13.12}$$

这里 $A\circ B$ 是 A 和 B 的 Hadamard 积.

证 对 n 进行归纳法.当 $n=1$ 时结论显然成立.

现在假设 $n\ge 2$,而且结论对阶不超过 $n-1$ 的所有矩阵成立,则有

$$(\det A_{11})\prod_{i=2}^{n}b_{ii}\le\det(A_{11}\circ B_{11}),$$

其中所用记号与 **10.13.11** 中的相同.注意,

$$A_{11}\circ B_{11}=(A\circ B)_{11},$$

且依 **10.13.11**, $A-\alpha(A)E_{11}$ 是半正定的;从而,依 **7.2.9**,

$$(A-\alpha(A)E_{11})\circ B$$

是半正定的.因此

$$\begin{aligned}0&\le\det[(A-\alpha(A)E_{11})\circ B]\\&=\det(A\circ B)-\alpha(A)b_{11}\det(A_{11}\circ B_{11}).\end{aligned}$$

推出

$$\begin{aligned}\det(A\circ B)&\ge\alpha(A)b_{11}\det(A_{11}\circ B_{11})\\&\ge\alpha(A)b_{11}(\det A_{11})\prod_{i=2}^{n}b_{ii}=(\det A)\prod_{i=1}^{n}b_{ii}.\end{aligned}$$ □

10.13.13 Ostrowski-Taussky 定理 设 $A\in\mathbb{C}^{n\times n}$ 使得

$$H(A) \equiv \frac{A+A^*}{2}$$

是正定矩阵,则

$$\det H(A) \le |\det A|, \tag{13.13}$$

等号成立当且仅当 A 为自伴矩阵.

证 令

$$S(A) \equiv \frac{A-A^*}{2}.$$

于是, $A = H(A) + S(A)$,不等式(13.13)等价于

$$\left|\det\left(I + H(A)^{-1} S(A)\right)\right| \ge 1.$$

然而 $H(A)^{-1} S(A)$ 相似于斜自伴矩阵

$$H(A)^{-1/2} S(A) H(A)^{-1/2}. \tag{13.14}$$

依 **3.4.2**,矩阵(13.14)从而 $H(A)^{-1} S(A)$ 只有纯虚数特征值.因此 $I + H(A)^{-1} S(A)$ 的特征值形如 $1 + \mathrm{i}t$,满足

$$|1 + \mathrm{i}t| \ge 1, \quad \forall t \in \mathbb{R}.$$

现设 $\mathrm{i}t_1, \cdots, \mathrm{i}t_n$ 是 $H(A)^{-1} S(A)$ 的特征值,则

$$\left|\det\left(I + H(A)^{-1} S(A)\right)\right| = \prod_{j=1}^{n} |1 + \mathrm{i}t_j| \ge 1.$$

而且,等号成立当且仅当 $t_1 = \cdots = t_n = 0$,等价于 $S(A) = 0$ (因为斜自伴矩阵是可对角化的). □

10.13.14 Minkowski 不等式 设 $A, B \in \mathbb{C}^{n\times n}$ 是正定矩阵,则

$$[\det(A+B)]^{1/n} \ge (\det A)^{1/n} + (\det B)^{1/n}. \tag{13.15}$$

证 设想对(13.15)同时左乘与右乘以 $(\det A^{-1/2})^{1/n}$,立即看出,不失一般性,可假设 $A = I$,故只须证明

$$[\det(I+B)]^{1/n} \ge 1 + (\det B)^{1/n}. \tag{13.16}$$

再进一步,如果$0<\lambda_1\le\cdots\le\lambda_n$是$B$的特征值,那么不等式(13.16)等价于

$$\prod_{j=1}^{n}\left(1+\lambda_j\right)\ge\left(1+\sqrt[n]{\lambda_1\cdots\lambda_n}\right)^n,$$

而此不等式可以利用算术-几何平均不等式,通过直接逐项比较其两边的展开式而加以证明. □

10.14 矩阵的值域和数值半径

10.14.1 定义 设映射$F(\cdot):\mathbb{C}^{n\times n}\to\mathbb{C}$为

$$F(A)\equiv\{x^*Ax:x\in\mathbb{C}^n,x^*x=1\},\ \forall A\in\mathbb{C}^{n\times n},\quad(14.1)$$

$F(A)$称为矩阵A的**值域**(field of values)或**数值范围**(numerical range).

10.14.2 定理 矩阵的值域有以下基本性质:

(1) $F(I)=\{1\};F(\alpha I)=\{\alpha\},\forall\alpha\in\mathbb{C}$.

(2) 紧性.设$A\in\mathbb{C}^{n\times n}$,则$F(A)$是$\mathbb{C}$的紧集.

(3) 凸性.设$A\in\mathbb{C}^{n\times n}$,则$F(A)$是$\mathbb{C}$的凸子集.

(4) 平移.$F(A+\alpha I)=F(A)+\alpha,\forall A\in\mathbb{C}^{n\times n},\alpha\in\mathbb{C}$.

(5) 数乘.$F(\alpha A)=\alpha F(A),\forall A\in\mathbb{C}^{n\times n},\alpha\in\mathbb{C}$.

(6) 投影.设$A\in\mathbb{C}^{n\times n}$,则

$$F(H(A))=\operatorname{Re}F(A),\ \forall A\in\mathbb{C}^{n\times n},$$

这里$H(A)\equiv\frac{1}{2}\left(A+A^*\right)$是$A$的自伴部分.

(7) 正定指标函数.设$A\in\mathbb{C}^{n\times n}$,则

$$F(A)\subset\{z\in\mathbb{C}:\operatorname{Re}z>0\}$$

的充分必要条件是 $A+A^*$ 为正定矩阵.

(8) 半正定指标函数.设 $A\in\mathbb{C}^{n\times n}$,则

$$F(A)\subset\{z\in\mathbb{C}:\operatorname{Re}z\geq 0\}$$

的充分必要条件是 $A+A^*$ 为半正定矩阵.

(9) 谱包含关系.设 $A\in\mathbb{C}^{n\times n}$,则

$$\lambda(A)\subset F(A).$$

(10) 次加性. $F(A+B)\subset F(A)+F(B),\forall A,B\in\mathbb{C}^{n\times n}$.

(11) 酉相似不变性.对任何酉矩阵 $U\in\mathbb{C}^{n\times n}$,

$$F(U^*AU)=F(A),\quad\forall A\in\mathbb{C}^{n\times n}.$$

(12) 正规性.设 $A\in\mathbb{C}^{n\times n}$ 是正规矩阵,则

$$F(A)=\mathrm{Co}(\lambda(A)).$$

这里 $\mathrm{Co}(\lambda(A))$ 表示 A 的特征值集合 $\lambda(A)$ 的凸包.

(13) 直和.成立

$$F(A\oplus B)=\mathrm{Co}(F(A)\cup F(B)),$$
$$\forall A\in\mathbb{C}^{n_1\times n_1},B\in\mathbb{C}^{n_2\times n_2}.$$

(14) 子矩阵包含关系.设 $A\in\mathbb{C}^{n\times n},\alpha\subset\{1,\cdots,n\}$ 是指标集,则有

$$F(A(\alpha))\subset F(A).$$

证 (1) 留作练习.

(2) 集 $F(A)$ 是域

$$\{x\in\mathbb{C}^n:x^*x=1\}\tag{14.2}$$

上连续函数 $x\mapsto x^*Ax$ 的值域.域(14.2)是 Euclid 单位球的球面,是紧集.因紧集在连续映射下的象是紧集,故 $F(A)$ 是紧集.

(3) 证明从略.可作练习.

(4) 根据(14.1),

$$\begin{aligned}F(A+\alpha I)&=\{x^*(A+\alpha I)x:x^*x=1\}\\&=\{x^*Ax+\alpha x^*x:x^*x=1\}=\{x^*Ax+\alpha:x^*x=1\}\end{aligned}$$

$$= \{x^* Ax : x^* x = 1\} + \alpha = F(A) + \alpha .$$

(5) 根据(14.1),

$$F(\alpha A) = \{x^*(\alpha A)x : x^* x = 1\} = \{\alpha x^* Ax : x^* x = 1\}$$
$$= \alpha\{x^* Ax : x^* x = 1\} = \alpha F(A).$$

(6) 由于

$$x^* H(A)x = x^*\left[\frac{1}{2}(A + A^*)\right]x = \frac{1}{2}(x^* Ax + x^* A^* x)$$
$$= \frac{1}{2}\left(x^* Ax + (x^* Ax)^*\right) = \frac{1}{2}\left(x^* Ax + \overline{x^* Ax}\right)$$
$$= \operatorname{Re} x^* Ax . \tag{14.3}$$

因此,$F(H(A))$中的每一点是某一$z \in F(A)$的实部$\operatorname{Re} z$,反之亦然.

(7) $F(A) \subset \{z \in \mathbb{C}: \operatorname{Re} z > 0\}$等价于

$$\operatorname{Re} x^* Ax > 0, \quad \forall x \in \mathbb{C}^n .$$

而这一条件,从(14.3)知,等价于$A + A^*$正定.

(8) 证明与(7)相仿,留作练习.

(9) 设$\lambda \in \lambda(A)$,并取对应的特征向量$x \in \mathbb{C}^n$为单位向量,$Ax = \lambda x$,则

$$\lambda = \lambda x^* x = x^*(\lambda x) = x^* Ax \in F(A).$$

(10) 根据(14.1),

$$F(A + B) = \{x^*(A + B)x : x^* x = 1\}$$
$$= \{x^* Ax + x^* Bx : x^* x = 1\}$$
$$\subset \{x^* Ax : x^* x = 1\} + \{y^* By : y^* y = 1\}$$
$$= F(A) + F(B).$$

(11) 对于$x \in \mathbb{C}^n$,$x^* x = 1$,有

$$x^*(U^* AU)x = y^* Ay \in F(A),$$

其中

$$y \equiv Ux, \quad y^* y = x^* U^* U x = x^* x = 1.$$

因此

$$F(U^* A U) \subset F(A).$$

应用这一包含关系,又得

$$\begin{aligned} F(A) &= F((UU^*)A(UU^*)) \\ &= F((U^*)^*(U^* A U)(U^*)) \subset F(U^* A U). \end{aligned}$$

(12) A 正规,依 **2.4.4**,存在酉矩阵 $U \in \mathbb{C}^{n\times n}$,使得

$$A = U^* \Lambda U, \quad \Lambda = \mathrm{diag}(\lambda_1, \cdots, \lambda_n).$$

利用(11),

$$\begin{aligned} F(A) &= F(\Lambda) \\ &= \left\{ x^* \Lambda x = \sum_{i=1}^{n} |x_i|^2 \lambda_i : x = (x_1, \cdots, x_n)^{\mathrm{T}}, x^* x = \sum_{i=1}^{n} |x_i|^2 = 1 \right\}, \end{aligned}$$

表明 $F(A)$ 是 A 的特征值的凸组合全体,亦即 $F(A) = \mathrm{Co}(\lambda(A))$.

(13) 注意 $A \oplus B \in \mathbb{C}^{(n_1+n_2)\times(n_1+n_2)}$.设 $v \in \mathbb{C}^{n_1+n_2}$ 是任一单位向量,将其划分为

$$v = \begin{bmatrix} x \\ y \end{bmatrix}, \qquad x \in \mathbb{C}^{n_1}, \ y \in \mathbb{C}^{n_2},$$

于是

$$v^*(A \oplus B)v = x^* A x + y^* B y.$$

如果 $y^* y = 1$,那么 $x = 0$,

$$v^*(A \oplus B)v = y^* B y \in F(B),$$

故 $F(A \oplus B) \supset F(B)$.同样,从 $x^* x = 1$ 又得 $F(A \oplus B) \supset F(A)$.因而

$$F(A \oplus B) \supset F(A) \cup F(B).$$

由此,鉴于 $F(A \oplus B)$ 是凸的,推出

$$F(A \oplus B) \supset \mathrm{Co}(F(A) \cup F(B)).$$

另一方面,当 $x^*x=0$ 即 $y^*y=1$ 时,

$$v^*(A\oplus B)v=y^*By\in F(B)\subset \mathrm{Co}(F(A)\cup F(B)).$$

而且当 $y^*y=0$ 即 $x^*x=1$ 时,

$$v^*(A\oplus B)v=x^*Ax\in F(A)\subset \mathrm{Co}(F(A)\cup F(B)).$$

现在假设 $v^*v=x^*x+y^*y=1$ 且 $x\neq 0, y\neq 0$,则有

$$v^*(A\oplus B)v=x^*Ax+y^*By=x^*x\left(\frac{x^*Ax}{x^*x}\right)+y^*y\left(\frac{y^*By}{y^*y}\right).$$

其最后的表示式是

$$\frac{x^*Ax}{x^*x}\in F(A) \quad 和 \quad \frac{y^*By}{y^*y}\in F(B)$$

的凸组合,必属于 $\mathrm{Co}(F(A)\cup F(B))$.因此,成立反包含关系

$$F(A\oplus B)\subset \mathrm{Co}(F(A)\cup F(B)).$$

(14) 设 $\alpha=\{j_1,\cdots,j_k\}, 1\le j_1\le\cdots\le j_k\le n$,并设

$$x=(x_1,\cdots,x_k)^{\mathrm T}\in\mathbb{C}^k, \quad x^*x=1.$$

在 x 中将零元素插入适当的位置,使之成为

$$\hat{x}=(\hat{x}_1,\cdots,\hat{x}_n)^{\mathrm T}\in\mathbb{C}^n,$$

其中

$$\hat{x}_{j_i}=x_i,\ i=1,\cdots,k;\quad \hat{x}_j=0,\ \forall j\notin\alpha.$$

于是,显然有

$$x^*A(\alpha)x=\hat{x}^*A\hat{x}, \quad \hat{x}^*\hat{x}=1,$$

这表明成立 $F(A(\alpha))\subset F(A)$. □

10.14.3 定义 设映射 $r(\bullet):\mathbb{C}^{n\times n}\to\mathbb{R}$ 为

$$r(A)\equiv\max\{|z|: z\in F(A)\},$$

$r(A)$ 称为矩阵 A 的**数值半径**(numerical radius).

10.14.4 定理 设 $A,B\in\mathbb{C}^{n\times n}$,成立下列性质:

(1) $r(\bullet)$ 具有正定性

$$r(A)>0, A\neq 0;\quad r(0)=0.$$

齐次性

$$r(cA)=|c|r(A),\ c\in\mathbb{C}.$$

次加性

$$r(A+B)\leq r(A)+r(B).$$

但不满足次乘性.

(2) $4r(\bullet)$具有次乘性,即

$$4r(AB)\leq[4r(A)][4r(B)],$$

4是具有如此性质的最小正常数.因此,$4r(\bullet)$是矩阵范数.

(3) $2r(\bullet)$对 Hadamard 积具有次乘性,即

$$2r(A\circ B)\leq[2r(A)][2r(B)],$$

2是具有如此性质的最小正常数.因此,$2r(\bullet)$是矩阵 Hadamard 次乘性范数.若A或B是正规的,则

$$r(A\circ B)\leq r(A)r(B).$$

(4) 幂不等式

$$r(A^m)\leq[r(A)]^m,\quad m=1,2,\cdots$$

(5) $\rho(A)\leq r(A)$.

(6) $r(A)=r(A^*)$.

(7) $\frac{1}{2}\|A\|_2\leq r(A)\leq\|A\|_2$.

(8) $r(\hat{A})\leq r(A)$,其中$\hat{A}$是A的任一主子矩阵.

(9) $r(A\oplus B)=\max\{r(A),r(B)\},\forall A\in\mathbb{C}^{n_1\times n_1},B\in\mathbb{C}^{n_2\times n_2}$.

(10) $r(A)\leq r(|A|)=\frac{1}{2}\rho\left(|A|+|A|^{\mathrm{T}}\right)$.

(11) 若$A\geq 0, H(A)=\frac{1}{2}\left(A+A^{\mathrm{T}}\right)$,则

$$r(A)=r(H(A))=\rho(H(A)).$$

而且

$$r(A)=\max\left\{x^{\mathrm{T}}Ax : x=(x_1,\cdots,x_n)^{\mathrm{T}}\geq 0, x^{\mathrm{T}}x=1\right\}.$$

(12) $\|A^m\|_2 \leq 2r(A)^m, m=1,2,\cdots$

定理证明从略.可作练习.

10.14.5 定义 $A\in\mathbb{C}^{n\times n}$ 称为**径向矩阵**(radial matrix),如果

$$\rho(A)=\|A\|_2.$$

10.14.6 定理 设 $A\in\mathbb{C}^{n\times n}$,则下列各条件等价:

(1) A 是径向矩阵,即成立 $\rho(A)=\|A\|_2$.

(2) A 的数值半径 $r(A)=\|A\|_2$.

(3) $\|A^n\|_2=(\|A\|_2)^n$.

(4) 对于不小于 A 的最小多项式次数的某个整数 m,成立

$$\|A^m\|_2=(\|A\|_2)^m.$$

(5) $\|A^k\|_2=(\|A\|_2)^k, k=1,2,\cdots$

(6) A 酉相似于

$$\|A\|_2(U\oplus B),$$

其中 $U\in\mathbb{C}^{m\times m}$ 是酉矩阵,$1\leq m\leq n$;$B\in\mathbb{C}^{(n-m)\times(n-m)}$,$\rho(B)<1$ 且 $\|B\|_2\leq 1$.

定理证明从略.

10.14.7 定理 设 $A\in\mathbb{C}^{n\times n}$,则

(1) 如果 A 是正规矩阵,那么 A 是径向矩阵.

(2) 如果 A 是径向矩阵,那么当 $n=2$ 时 A 是正规矩阵,而当 $n\geq 3$ 时 A 不一定是正规矩阵.

(3) 如果 A 是径向矩阵,那么 A^k 是径向矩阵,$k=1,2,\cdots$,但是,存在这样的矩阵,它的某正整数次幂是径向矩阵,而其本身不是径向矩阵.

定理证明从略.

10.14.8 定义 $A \in \mathbb{C}^{n\times n}$ 称为**谱矩阵**(spectral matrix),如果

$$\rho(A) = r(A).$$

10.14.9 定理 设 $A \in \mathbb{C}^{n\times n}$.

(1) 若 A 是径向矩阵,则 A 是谱矩阵.

(2) 若 A 是谱矩阵,λ 是 A 的特征值,$|\lambda| = \rho(A)$,则 λ 是 A 的正规特征值.

(3) 若 A 是正规矩阵,则 A 是谱矩阵.反之,若 A 是谱矩阵,则仅当 $n=2$ 时 A 是正规矩阵.

定理证明从略.

10.15 区间矩阵

10.15.1 定义 $n\times n$ 矩阵 $A=[a_{ij}]$ 称为**区间矩阵**(interval matrix),如果其每个元素是一个闭区间上的参变量,

$$\alpha_{ij} \le a_{ij} \le \beta_{ij}, \quad \alpha_{ij}, \beta_{ij} \in \mathbb{R}, \quad i,j=1,\cdots,n.$$

或者说,每个元素是一个闭区间,并记作

$$a_{ij} \equiv [\alpha_{ij}, \beta_{ij}] \subset \mathbb{R}, \quad i,j=1,\cdots,n.$$

区间矩阵在应用中的重要性在不断增长.

10.15.2 定义 设 $a \equiv [a_1, a_2]$, $b \equiv [b_1, b_2] \subset \mathbb{R}$ 是任意的闭区间.

a 与 b 称为**相等**,记作 $a=b$,如果

$$a_1 = b_1, \qquad a_2 = b_2.$$

a 加 b 之**和**是指

$$a+b \equiv [a_1+b_1, a_2+b_2].$$

a 减 b 之差是指

$$a-b\equiv[a_1-b_2,a_2-b_1].$$

a 乘 b 之积是指

$$a*b\equiv[\min c,\max c],\quad c=\{a_1b_1,a_1b_2,a_2b_1,a_2b_2\}.$$

10.15.3 定义 设 $A=[a_{ij}]$ 和 $B=[b_{ij}]$ 是任意的 $n\times n$ 区间矩阵.依 **10.15.2**,按通常矩阵的相等、加法与乘法,定义 A 与 B 的**相等**、**和**、**差**、**积**为

$$A=B\Leftrightarrow a_{ij}=b_{ij},1\le i,j\le n,$$

$$A\pm B=[a_{ij}]\pm[b_{ij}]=[a_{ij}\pm b_{ij}],$$

$$AB=[a_{ij}][b_{ij}]=\left[\sum_{k=1}^{n}a_{ik}*b_{kj}\right].$$

例 考虑区间矩阵

$$A=\begin{bmatrix}[1,3] & [-1,0]\\ [-2,2] & [3,6]\end{bmatrix},\quad B=\begin{bmatrix}[-2,-1] & [-3,1]\\ [0,4] & [-2,4]\end{bmatrix}. \tag{15.1}$$

依 **10.15.3** 和 **10.15.2**,

$$\begin{aligned}A+B&=\begin{bmatrix}[-1,2] & [-4,1]\\ [-2,6] & [1,10]\end{bmatrix},\\ A-B&=\begin{bmatrix}[2,5] & [-2,3]\\ [-6,2] & [-1,8]\end{bmatrix}.\end{aligned} \tag{15.2}$$

由于

$$\begin{aligned}a_{11}*b_{11}&=[1,3]*[-2,-1]\\ &=[\min\{-2,-1,-6,-3\},\max\{-2,-1,-6,-3\}]\\ &=[-6,-1],\\ a_{12}*b_{21}&=[-1,0]*[0,4]\end{aligned}$$

$$=[\min\{0,-4,0,0\},\max\{0,-4,0,0\}]$$
$$=[-4,0].$$

故

$$(AB)_{11}=a_{11}*b_{11}+a_{12}*b_{21}$$
$$=[-6,-1]+[-4,0]=[-10,-1].$$

类似地可算出 AB 的其它元素,最后得

$$AB=\begin{bmatrix}[-10,-1] & [-13,5]\\ [-4,28] & [-18,30]\end{bmatrix}. \tag{15.3}$$

10.15.4 定义 设 $a\equiv[a_1,a_2]\subset\mathbb{R}$是任意的闭区间,称

$$|a|\equiv\max\{|a_1|,|a_2|\}$$

为 a 的**绝对值**.称

$$d(a)\equiv a_2-a_1$$

为 a 的**宽度**.

显然,$|a|\geq 0$,$d(a)\geq 0$.

10.15.5定义 设 $a\equiv[a_1,a_2]$,$b\equiv[b_1,b_2]\subset\mathbb{R}$是任意的两个闭区间,称

$$q(a,b)\equiv\max\{|a_1-b_1|,|a_2-b_2|\}$$

为 a 和 b 的**距离**.

当 $a_1=a_2$ 时,区间 $a=[a_1,a_1]$称为**点区间**.

对于两个点区间,依距离定义,归结为两个实数的距离,即

$$q([a_1,a_1],[b_1,b_1])=|a_1-b_1|.$$

利用 **10.15.4** 和 **10.15.5**,可以构造如下三种非负矩阵.

10.15.6 定义 设 $A=[a_{ij}]$ 和 $B=[b_{ij}]$ 是任意 $n\times n$ 区间矩阵,称

$$|A|\equiv[\,|a_{ij}|\,]$$

为 A 的**绝对值矩阵**.称

$$d(A)\equiv[d(a_{ij})]$$

为 A 的**宽度矩阵**(width matrix).称

$$q(A,B)\equiv[q(a_{ij},b_{ij})]$$

为 A 和 B 的**距离矩阵**(distance matrix).

10.15.7 定理 设 A,B,C 是任意的 $n\times n$ 区间矩阵,则成立如下基本性质:

(1) $|A+B|\le|A|+|B|$.

(2) $|AB|\le|A||B|$.

(3) $d(A\pm B)\le d(A)+d(B)$.

(4) $|A|d(B)\le d(AB)\le d(A)|B|+|A|d(B)$.

(5) $q(A,B)\le q(A,C)+q(B,C)$.

(6) $q(AB,AC)\le|A|q(B,C)$.

证明留作练习.

例 对于(15.1)中区间矩阵,有

$$|A|=\begin{bmatrix}3&1\\2&6\end{bmatrix};\quad |B|=\begin{bmatrix}2&3\\4&4\end{bmatrix}$$

和

$$|A+B|=\begin{bmatrix}2&4\\6&10\end{bmatrix},\quad |A||B|=\begin{bmatrix}10&13\\28&30\end{bmatrix}.$$

从(15.3),

$$|AB|=\begin{bmatrix}10&13\\28&30\end{bmatrix},$$

故就此例来说,

$$|A+B|<|A|+|B|,\quad |AB|=|A||B|.$$

另外,

$$d(A)=\begin{bmatrix}2 & 1\\ 4 & 3\end{bmatrix},\quad d(B)=\begin{bmatrix}1 & 4\\ 4 & 6\end{bmatrix}.$$

从(15.2),

$$d(A+B)=\begin{bmatrix}3 & 5\\ 8 & 9\end{bmatrix}=d(A-B).$$

再从(15.3),

$$d(AB)=\begin{bmatrix}9 & 18\\ 32 & 48\end{bmatrix}.$$

现在,**10.15.7** 的(4)中的其它两个表示式是

$$|A|d(B)=\begin{bmatrix}7 & 18\\ 26 & 44\end{bmatrix}$$

和

$$d(A)|B|+|A|d(B)=\begin{bmatrix}8 & 10\\ 20 & 24\end{bmatrix}+\begin{bmatrix}7 & 18\\ 26 & 44\end{bmatrix}=\begin{bmatrix}15 & 28\\ 46 & 68\end{bmatrix},$$

易见满足 **10.15.7** 的(4).

关于区间矩阵的进一步内容可参看[AH1983].

10.16 若干特性矩阵

10.16.1 定义 设 $A=[a_{ij}]\in\mathbb{C}^{m\times n}$,而且

$$[A \quad B] = \begin{bmatrix} a_{11} & \cdots & a_{1n} & b_{11} & \cdots & b_{1l} \\ \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \\ a_{m1} & \cdots & a_{mn} & b_{m1} & \cdots & b_{ml} \end{bmatrix}, \qquad (16.1)$$

其中

$$B \equiv [b_{ij}] \in \mathbb{C}^{m \times l}.$$

$m \times (n+l)$矩阵

$$C \equiv [A \quad B]$$

称为 A 的**增广矩阵**(augmented matrix).

在讨论线性方程(组)的解时常用增广矩阵一词.

一般地说,设 $A = [a_{ij}] \in \mathbb{C}^{m \times n}$ 和 $B \equiv [b_{ij}] \in \mathbb{C}^{m \times l}$ 是已知矩阵,则矩阵线性方程

$$AX = B \qquad (16.2)$$

是否存在解——满足(16.2)的 $X = [x_{ij}] \in \mathbb{C}^{n \times l}$,假若存在又有多少解等,自然完全取决于 A 的增广矩阵 $C \equiv [A \quad B]$的性质.

矩阵线性方程(16.2)等价于

$$Ax^{(j)} = b^{(j)}, \quad j = 1, \cdots, l, \qquad (16.3)$$

其中

$$x^{(j)} \equiv \begin{pmatrix} x_{1j} \\ \vdots \\ x_{nj} \end{pmatrix}, \quad b^{(j)} \equiv \begin{pmatrix} b_{1j} \\ \vdots \\ b_{mj} \end{pmatrix}, \quad j = 1, \cdots, l.$$

于是,只须讨论具有 n 个未知数 m 个方程的线性方程组

$$Ax = b, \qquad (16.4)$$

其中

$$x = (x_1, \cdots, x_n)^{\mathrm{T}}, \quad b = (b_1, \cdots, b_m)^{\mathrm{T}}.$$

此时,增广矩阵

$$C \equiv [A \quad b] \qquad (16.5)$$

是 $m\times(n+1)$ 矩阵.

10.16.2 定理 设 $A\in\mathbb{C}^{m\times n}, b\in\mathbb{C}^{m}, C\equiv[A\quad b]$ 是 A 的增广矩阵,则

$$\operatorname{rank}A\le\operatorname{rank}C.$$

而且对于线性方程组(16.4),成立

(1) *存在唯一解的充分必要条件是*

$$\operatorname{rank}A=\operatorname{rank}C=n.$$

(2) *存在无穷多个解的充分必要条件是*

$$\operatorname{rank}A=\operatorname{rank}C<n.$$

(3) 无解的充分必要条件是

$$\operatorname{rank}A<\operatorname{rank}C.$$

定理证明从略. 这是线性代数中的基础结论.可参看[M1963].其证明还可在许多线性代数教科书中找到.

10.16.3 定义 $A\in\mathbb{C}^{n\times n}$ 称为**正交随机矩阵**(orthostochastic matrix),如果存在酉矩阵 $U\in\mathbb{C}^{n\times n}$,使得

$$A=U\circ\overline{U}, \tag{16.6}$$

这里 $U\circ\overline{U}$ 是 U 和 $\overline{U}$ 的 Hadamard 积.

10.16.4 定理 每个正交随机矩阵是双随机矩阵.反之则不然.

证 注意到酉矩阵的各行和各列都是单位向量,即可从正交随机矩阵推出其必为双随机矩阵.

考虑矩阵

$$A=\begin{bmatrix}\frac{1}{2} & 0 & \frac{1}{2}\\ \frac{1}{2} & \frac{1}{2} & 0\\ 0 & \frac{1}{2} & \frac{1}{2}\end{bmatrix}, \tag{16.7}$$

A 显然是双随机矩阵.

但是, A 不是正交随机矩阵(留作练习). □

10.16.5 定理 设 $A \in \mathbb{C}^{n\times n}$ 是正交随机矩阵,则

$$\|A\|_2 = 1,$$

而且 A 是径向矩阵.反之则不然.

证 依 **10.16.4**, A 是双随机矩阵,再依 **4.5.3**,有

$$\rho(A) = 1, \quad \|A\|_2 \ge 1.$$

另一方面,设

$$A = U \circ \overline{U},$$

U 是酉矩阵.依 **7.2.16**,又有

$$\|A\|_2 = \|U \circ \overline{U}\|_2 = \sigma_1(U \circ \overline{U}) \le \sigma_1(U)\sigma_1(\overline{U}) = 1,$$

因此

$$\|A\|_2 = \rho(A) = 1,$$

A 是径向矩阵.

反过来的结论是不成立的.

(16.7)的矩阵 A 是径向矩阵,而且 $\|A\|_2 = 1$,但不是正交随机矩阵. □

10.16.6 定义 $A \in \mathbb{C}^{n\times n}$ 称为**相关矩阵**(correlation matrix),如果 A 是半正定的,而且其所有对角元素等于1.

10.16.7 定理 设 $A = [a_{ij}] \in \mathbb{C}^{n\times n}$ 是正定矩阵,并设

$$P \equiv \operatorname{diag}\left(\frac{1}{\sqrt{a_{11}}}, \cdots, \frac{1}{\sqrt{a_{nn}}}\right),$$

则有

$$PAP=\left[a_{ij}/\sqrt{a_{ii}a_{jj}}\right]$$

是正定的,而且是相关矩阵,并满足

$$\left|\frac{a_{ij}}{\sqrt{a_{ii}a_{jj}}}\right|<1,\quad i\neq j,\ i,j=1,\cdots,n.$$

证 因为 A 正定,对于任何 $x\in\mathbb{C}^n$, $x\neq 0$,显然有

$$(PAPx,x)=(A(Px),(Px))>0,$$

所以 PAP 是正定的.而且, PAP 的对角元素全为1,故 PAP 是相关矩阵.

依 **3.10.6**,正定矩阵的任一主子式大于零.因此,考虑 PAP 的2阶主子式,注意到 $a_{ji}=a_{ij}^*$,得

$$0<\begin{vmatrix}1 & \dfrac{a_{ij}}{\sqrt{a_{ii}a_{jj}}}\\ \dfrac{a_{ji}}{\sqrt{a_{ii}a_{jj}}} & 1\end{vmatrix}=1-\left(\frac{|a_{ij}|}{\sqrt{a_{ii}a_{jj}}}\right)^2,$$

$$i\neq j,\ i,j=1,\cdots,n.$$ □

10.16.8 推论 设 $A=[a_{ij}]\in\mathbb{C}^{n\times n}$ 是相关矩阵,则

$$|a_{ij}|<1,\quad i\neq j,\ i,j=1,\cdots,n.$$

而且,当 A 正定时成立严格不等号. □

10.16.9 定义 $A\in\mathbb{R}^{n\times n}$ 称为**单调矩阵**(monotone matrix),如果

$$x\in\mathbb{R}^n,\ Ax\geq 0\Rightarrow x\geq 0. \tag{16.8}$$

由定义直接推出:如果 A 是单调矩阵,则

$$x,y\in\mathbb{R}^n,\ Ax\geq Ay\Rightarrow x\geq y. \tag{16.9}$$

10.16.10 定理 $A \in \mathbb{R}^{n\times n}$ 为单调矩阵的充分必要条件是 A 非奇异,而且 $A^{-1} \geq 0$.

证 如果 $A^{-1} \geq 0$,那么从

$$Ax \equiv y \geq 0$$

即可推出

$$x = A^{-1}y \geq 0.$$

现在,假设 A 单调.此时,若 x 满足

$$Ax = 0, \tag{16.10}$$

则同时有 $A(-x)=0$.从 A 单调,有

$$Ax = 0 \Rightarrow x \geq 0$$

和

$$A(-x)=0 \Rightarrow -x \geq 0,$$

即有 $x = 0$.这表明齐次方程组(16.10)只有零解,等价于 A 是非奇异矩阵.

因此,对于 $j = 1,\cdots,n$,方程组

$$Ax = e_j$$

必有唯一解,记其为 $x^{(j)}$;这里 e_j 是标准单位向量,即单位矩阵 I_n 的第 j 列.

然后,仍由 A 单调,对于 $j = 1,\cdots,n$,得知

$$x^{(j)} \geq 0,$$

从而

$$A^{-1}e_j = x^{(j)} \geq 0.$$

这就是说,矩阵 A^{-1} 的第 j 列是非负的.

于是,$A^{-1} \geq 0$. □

10.16.11 定理 设 $A \in \mathbb{R}^{n\times n}$ 的非对角元素是非正的,则 A 为单调矩阵的充分必要条件是 A 为 M-矩阵.

这一定理是 **4.6.5** 中的(1)$\Leftrightarrow$(2)的又一表述.

例 考虑 $n\times n$ 矩阵

$$A=\begin{bmatrix}1 & & & & \\ -2 & \ddots & & \text{\huge 0} & \\ 1 & \ddots & \ddots & & \\ & \ddots & \ddots & \ddots & \\ \text{\huge 0} & & 1 & -2 & 1\end{bmatrix}$$

和

$$B=\begin{bmatrix}2 & -1 & & \text{\huge 0} & \\ -1 & \ddots & \ddots & & \\ & \ddots & \ddots & \ddots & \\ \text{\huge 0} & & \ddots & 2 & -1 \\ & & & -1 & 1\end{bmatrix}.$$

由于

$$A=C^2,\qquad B=C^{\mathrm{T}}C,$$

其中

$$C=\begin{bmatrix}1 & & \text{\huge 0} & \\ -1 & \ddots & & \\ & \ddots & \ddots & \\ \text{\huge 0} & & -1 & 1\end{bmatrix},$$

容易算出

$$C^{-1}=\begin{bmatrix}1 & & \text{\huge 0} & \\ \vdots & \ddots & & \\ \vdots & & \ddots & \\ 1 & \cdots & \cdots & 1\end{bmatrix}.$$

因此,有

$$A^{-1}=\left(C^{-1}\right)^2=\begin{bmatrix}1 & & & & & \\ 2 & \ddots & & 0 & & \\ 3 & \ddots & \ddots & & & \\ \vdots & \ddots & \ddots & \ddots & & \\ n-1 & & \ddots & \ddots & \ddots & \\ n & n-1 & \cdots & 3 & 2 & 1\end{bmatrix}$$

和

$$B^{-1}=C^{-1}\left(C^{-1}\right)^{\mathrm{T}}=\begin{bmatrix}1 & 1 & 1 & \cdots & 1 & 1\\ 1 & 2 & 2 & \cdots & 2 & 2\\ 1 & 2 & 3 & \cdots & 3 & 3\\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots\\ 1 & 2 & 3 & \cdots & n-1 & n-1\\ 1 & 2 & 3 & \cdots & n-1 & n\end{bmatrix},$$

所以,A 和 B 都是单调矩阵,而且 B 是 M-矩阵.

10.16.12 定理 设 $A\in\mathbb{R}^{n\times n}$ 是单调矩阵,而且其非对角元素是非正的.又设 $y,z\in\mathbb{R}^n$ 满足

$$Ay\geq b,\qquad Az\leq b,$$

则有

$$y\geq x\geq z,$$

其中 x 是方程组 $Ax=b$ 的唯一解.

证明留作练习.

10.16.13 定义 设 $A\in\mathbb{C}^{m\times n}$,且谱范数 $\|A\|_2\leq 1$,则称 A 为**压缩矩阵**(contraction matrix);特别,当 $\|A\|_2<1$ 时,则称 A 为**严格压缩矩阵**.

10.16.14 定理　设 $A,B\in\mathbb{C}^{n\times n}$ 是压缩矩阵,则 AB 是压缩矩阵.

证　由 $\|AB\|_2\le\|A\|_2\|B\|_2$ 即得结论. □

一般,任何有限个压缩矩阵之积是压缩矩阵.

10.16.15 定义　设 $A\in\mathbb{C}^{n\times n}$.如果 $x\in\mathbb{C}^n$ 使得 $Ax=x$,则称向量 x 是 A 的**不动点**.

定义表明,如果 A 有不动点 $x\ne 0$,那么 A 有特征值 $\lambda=1$,而且 x 就是对应的特征向量.

10.16.16 定理　设 $A\in\mathbb{C}^{n\times n}$ 是压缩矩阵,则 A 的每个不动点也是 A^* 的不动点.

证　注意到

$$\|A^*\|_2=\|A\|_2\le 1,$$

A^* 也是压缩矩阵.

现在设 x 是 A 的不动点.因为 $Ax=x$,

$$\begin{aligned}\|A^*x-x\|_2^2&=\|A^*Ax-x\|_2^2=\|A^*Ax\|_2^2-2\|Ax\|_2^2+\|x\|_2^2\\&=\|A^*Ax\|_2^2-\|Ax\|_2^2\le\|Ax\|_2^2-\|Ax\|_2^2=0,\end{aligned}$$

所以 $A^*x=x$. □

当 A 不是压缩矩阵时,以上定理的结论一般不成立.例如

$$A=\begin{bmatrix}1&1\\0&0\end{bmatrix},\qquad x=\begin{bmatrix}1\\0\end{bmatrix},$$

容易验证 A 不是压缩矩阵(留作练习).显然,满足 $Ax=x$,x 是 A 的不动点.但是

$$A^*=\begin{bmatrix}1&0\\1&0\end{bmatrix},\qquad \begin{bmatrix}1&0\\1&0\end{bmatrix}\begin{bmatrix}1\\0\end{bmatrix}=\begin{bmatrix}1\\1\end{bmatrix},$$

x 不是 A^* 的不动点.

10.16.17 定理 $A\in\mathbb{C}^{n\times n}$ 为压缩矩阵的充分必要条件是 A 为 $\mathbb{C}^{n\times n}$ 中酉矩阵的有限凸组合.

证 必要性.依 **5.4.10** 的(2), A 可表示为

$$A=\mu_1K_1+\cdots+\mu_nK_n, \tag{16.11}$$

其中

$$K_i=UE_iV^*,\ E_i=(e_1,\cdots,e_i,0,\cdots,0),\ i=1,\cdots,n,$$
$$\mu_i=\sigma_i-\sigma_{i+1},\ i=1,\cdots,n-1;\quad \mu_n=\sigma_n.$$

这里 $U,V\in\mathbb{C}^{n\times n}$ 是酉矩阵, $\sigma_1\ge\cdots\ge\sigma_n\ge 0$ 是 A 的奇异值.

因为 A 是压缩矩阵,所以

$$\mu_1+\cdots+\mu_n=\sigma_1=\rho(A^*A)^{1/2}=\|A\|_2\le 1.$$

注意到

$$E_i=\frac{1}{2}(I+\hat{E}_i);$$
$$\hat{E}_i\equiv(e_1,\cdots,e_i,-e_{i+1},\cdots,-e_n),\ i=1,\cdots,n.$$

于是

$$K_i=\frac{1}{2}UV^*+\frac{1}{2}U\hat{E}_iV^*,\ i=1,\cdots,n,$$

其中 UV^* 是酉矩阵,而且由于

$$(U\hat{E}_iV^*)^*(U\hat{E}_iV^*)=V\hat{E}_iU^*U\hat{E}_iV^*=I,$$

因此每个 $U\hat{E}_iV^*$ 也是酉矩阵.这样,(16.11)的右端就是酉矩阵的有限凸组合.

充分性.设 A 的酉矩阵的有限凸组合形如(16.11),则因系数 $\mu_1,\cdots,\mu_n\ge 0$,

$$\|A\|_2 \le \mu_1 \|K_1\|_2 + \cdots + \mu_n \|K_n\|_2 = \mu_1 + \cdots + \mu_n = 1,$$

从而 A 是压缩矩阵. □

10.16.18 定理 设 $L \in \mathbb{C}^{m\times m}, M \in \mathbb{C}^{n\times n}, X \in \mathbb{C}^{m\times n}$,矩阵

$$\begin{bmatrix} L & X \\ X^* & M \end{bmatrix} \in \mathbb{C}^{(m+n)\times(m+n)} \tag{16.12}$$

为半正定的充分必要条件是 L 和 M 为半正定的,而且存在压缩矩阵 $C \in \mathbb{C}^{m\times n}$ 使得

$$X = L^{1/2} C M^{1/2}.$$

证 先假设 L 和 M 是正定的.依 **10.13.5**,块矩阵(16.12)半正定的充分必要条件是

$$\begin{aligned} 1 \ge \rho\left(X^* L^{-1} X M^{-1}\right) &= \rho\left(M^{-1/2} X^* L^{-1} X M^{-1/2}\right) \\ &= \rho\left(\left(L^{-1/2} X M^{-1/2}\right)^* \left(L^{-1/2} X M^{-1/2}\right)\right) \\ &= \left(\sigma_1\left(L^{-1/2} X M^{-1/2}\right)\right)^2. \end{aligned} \tag{16.13}$$

令

$$C = L^{-1/2} X M^{-1/2},$$

故 $X = L^{1/2} C M^{1/2}$.(16.13)表明 C 是压缩矩阵.

对于一般 L 和 M 是半正定的情形,通过极限论证方法即可推出. □

对于给定的 $X \in \mathbb{C}^{m\times n}$,存在 L 和 M 的一些有用的选取,使得形如(16.12)的块矩阵是半正定的.例如:

(1) 若

$$X = Y^* Z, \qquad Y \in \mathbb{C}^{r\times m},\ Z \in \mathbb{C}^{r\times n},$$

则

$$\begin{bmatrix} Y^*Y & X \\ X^* & Z^*Z \end{bmatrix} = \begin{bmatrix} Y^*Y & Y^*Z \\ Z^*Y & Z^*Z \end{bmatrix} = \begin{bmatrix} Y & Z \end{bmatrix}^* \begin{bmatrix} Y & Z \end{bmatrix}$$

显然是半正定的.

(2) 10.16.18 蕴涵

$$\begin{bmatrix} \sigma_1(X)I & X \\ X^* & \sigma_1(X)I \end{bmatrix}$$

是半正定的,因为$(\sigma_1(X))^{-1}X$必是压缩矩阵.

10.16.19 定义 设$A\in\mathbb{C}^{n\times n}$,$P_1,P_2,\cdots,P_{n!}\in\mathbb{C}^{n\times n}$是全部$n!$个按确定顺序排列的置换矩阵.

定义矩阵

$$\Pi(\cdot)\in\mathbb{C}^{n!\times n!},$$

$\Pi(A)$的(i,j)-元素是$P_i^{\mathrm{T}}AP_j$的主对角元素之积. $\Pi(A)$称为A的 **Soules 矩阵**.

注意,$\Pi(\cdot)$的每个主对角元素是A的主对角元素之积.

10.16.20 定理 设$A\in\mathbb{C}^{n\times n}$,$\Pi(\cdot)$是$A$的 Soules 矩阵,则

(1) 如果A是自伴矩阵,那么$\Pi(A)$也是自伴矩阵.

(2) 如果A是正定矩阵,那么$\Pi(A)$也是正定矩阵.

(3) $\det A$是$\Pi(A)$的特征值.

(4) $\mathrm{per}A\equiv\sum_p a_{p_1 1}\cdots a_{p_n n}$是$\Pi(A)$的特征值,这里求和是对$1,\cdots,n$的所有置换$p$而言的,

$$p(1,\cdots,n)=(p_1,\cdots,p_n).$$

“per”是 permanent 的缩写.

定理证明从略.

10.16.21 定义 $A=[a_{ij}]\in\mathbb{R}^{n\times n}$称为$\hat{\mathrm{M}}$-**矩阵**,如果满足

(1) $a_{ii}>0$, $i=1,\cdots,n-1$; $a_{nn}\geq 0$;

(2) $a_{ij} \le 0$, $i \ne j$, $i,j = 1,\cdots,n$;

(3) $n(i) > i$, $i = 1,\cdots,n-1$; 其中

$$n(i) = \max\{j : 1 \le j \le n, a_{ij} \ne 0\},$$

即为 A 的第 i 行元素中最后一个非零元素所在的列.

$\hat{\text{M}}$-矩阵有关不完全LU分解计算的结论可参看[X1995].

10.17 某些应用矩阵

10.17.1 定义 $A = [a_{ij}] \in \mathbb{R}^{n\times n}$ 称为**本质非负矩阵**(essentially nonnegative matrix),如果其所有非对角元素是非负的,即

$$a_{ij} \ge 0, \quad i \ne j, \ i,j = 1,\cdots,n.$$

10.17.2 定理 设 $A \in \mathbb{R}^{n\times n}$ 是本质非负矩阵,则

(1) 存在 $\lambda \in \mathbb{R}, \lambda > 0$,使得 $\lambda I + A \ge 0$.

(2) A 有一个实的特征值 $r(A)$,满足

$$r(A) \ge \mathrm{Re}\,\lambda_i, \quad \forall \lambda_i \in \lambda(A),$$

$r(A)$ 称为 A 的**优势特征值**(dominant eigenvalue).

(3) $r(A)$ 不一定是 A 的模最大的特征值,但当 $A \ge 0$ 时,有 $r(A) = \rho(A)$.

证明留作练习.提示:利用矩阵 $\lambda I + A$ 的特征值是

$$\lambda + \lambda_i, \quad \forall \lambda_i \in \lambda(A)$$

10.17.3 定理 设 $A \in \mathbb{R}^{n\times n}$ 是本质非负矩阵,则矩阵

$$A + D, \quad \forall D = \mathrm{diag}(x_1,\cdots,x_n) \in \mathbb{R}^{n\times n}$$

是本质非负的.而且,优势特征值$r(A+D)$是$x=(x_1,\cdots,x_n)^T$的凸函数.

定理证明从略.可作练习.

10.17.4 定义 $A=[a_{ij}]\in\mathbb{R}^{n\times n}$称为**本质正矩阵**(essentially positive matrix),如果

$$a_{ij}\ge 0,\quad i\ne j,\ i,j=1,\cdots,n,$$

而且A是不可约的.

本质正矩阵出现在经济学研究的问题中,也称为**输入-输出矩阵**(input-output matrix)和 **Leon*tie*ff 矩阵**.

从定义直接推出:A为本质正矩阵的充分必要条件是存在实数α,$\alpha I+A$是非负的、不可约的,而且是素矩阵.

10.17.5 定理 $A=[a_{ij}]\in\mathbb{R}^{n\times n}$为本质正矩阵的充分必要条件是

$$e^{tA}>0,\quad \forall t>0. \tag{17.1}$$

证 假设成立

$$e^{tA}=I+tA+\frac{t^2}{2!}A^2+\cdots>0,\quad \forall t>0, \tag{17.2}$$

则A必是不可约的.不然的话,A及其所有幂是可约的,蕴涵e^{tA}有若干零元素,矛盾.再者,如果存在某$a_{ij}<0,i\ne j$,那么从(17.2)看出,对于充分小的$t>0$,$(e^{tA})_{ij}<0$,矛盾.因此,A是本质正矩阵.

现在假设A是本质正矩阵.那么,对于充分大的实数$\alpha>0$,$\alpha I+A$是非负的、不可约的,而且是素矩阵.因为$\alpha I+A$的幂都是非负的,而且依**4.3.9**,其所有充分高次幂全是正的,所以

$$e^{\alpha I+A}>0.$$

于是,容易验证

$$e^{tA}=e^{-\alpha tI}e^{t(\alpha I+A)}>0,\quad \forall t>0. \qquad \square$$

10.17.6 定理 设 A 是本质正矩阵,则 A 有一实特征值 $r(A)$ 满足

(1) $r(A)$ 对应的特征向量可取作正的.

(2) $\operatorname{Re}\lambda<r(A), \forall\lambda\in\lambda(A)$.

(3) $r(A)$ 随 A 的任一元素递增而递增.

(4) $r(A)$ 是单特征值.

证 由于 A 是本质正的,存在实数 α, $B\equiv\alpha I+A$ 是非负的,不可约的,而且是素矩阵.依 **4.2.4**,存在向量 $x>0$,使得

$$Bx=\rho(B)x,$$

因此

$$Ax=(\rho(B)-\alpha)x\equiv r(A)x,$$

结论(2)和(4)类似地可以从 **4.2.4** 推出.(3)的证明从略. $\square$

10.17.7 定义 设 A 是本质正矩阵,$r(A)$ 是 **10.17.6** 中所述的 A 的实特征值.按照 $r(A)$ 为正的、零和负的,分别称矩阵 A 是**超临界**的(supercritical)、**临界**的(critical)和**亚临界**的(subcritcal).

这些术语出自于核反应堆理论的有关数学模型的研究.

可以证明:当 A 是本质正矩阵时,成立

$$\left\|e^{tA}\right\|\sim Ke^{tr(A)},\quad t\to\infty, \tag{17.3}$$

其中 K 是正常数.(17.3)表明对于大的 t, $r(A)$ 支配着 $\left\|e^{tA}\right\|$ 的渐近性态.

10.17.8 定理 A 为本质非负矩阵的充分必要条件是

$$e^{tA}\geq 0,\quad \forall t\geq 0.$$

证明可利用 **10.17.5**,留作练习.

10.17.9 定义 $H_n \in \mathbb{R}^{n\times n}$ 称为 ***Hadamard* 矩阵**,如果它的所有元素为 $+1$ 或 -1,而且满足

$$H_n H_n^{\mathrm{T}} = H_n^{\mathrm{T}} H_n = nI_n. \tag{17.4}$$

容易看出,用 -1 乘 H_n 的任一行或任一列的元素,所得之积是又一个 Hadamard 矩阵.如此,从每个 Hadamard 矩阵 H_n 出发,可以产生第1行和第1列所有元素全为 $+1$ 的 Hadamard 矩阵,并称之为是**规范化**的.

从(17.4)直接推出:$\sqrt{n}H_n$ 是正交矩阵.

10.17.10 定理 仅当 $n=1$,或 2,或 4 的倍数时存在 Hadamard 矩阵 H_n.

证 设 $H_n = [h_{ij}] \in \mathbb{R}^{n\times n}$, $h_{ij} = +1$ 或 -1, $\forall i, j = 1,\cdots,n$.

当 $n=1$ 时, H_1 显然是 Hadamard 矩阵.

当 $n=2$ 时,

$$H_2 = \begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix} \tag{17.5}$$

是 Hadamard 矩阵,而且还是规范化的.事实上,

$$H_2 H_2^{\mathrm{T}} = H_2^{\mathrm{T}} H_2 = (H_2)^2 = \begin{bmatrix} 2 & 0 \\ 0 & 2 \end{bmatrix} = 2I_2.$$

现设 $n \geq 3$.如果 H_n 是 Hadamard 矩阵,那么因行与行之间的正交性及元素全为 $+1$ 或 -1, H_n 必须满足:

(1) 第2行正好有一半即 $n/2$ 个元素与第1行的对应元素反号;

(2) 第3行正好有 $n/2$ 个元素与第1行的对应元素反号,又正好

有$n/2$个元素与第2行的对应元素反号.

由(1)推出$n/2$应为整数,即n为偶数;再由(2)推出$(n/2)/2$应为整数,即n为4的倍数.这样,当$n \ge 3$时存在 Hadamard 矩阵的必要条件是n为4的倍数.

反过来还须证明n为4的倍数时必存在Hadamard矩阵.证明从略. □

这里指出,利用(17.5)的H_2及 Kronecker 积可以构造一族规范化对称 Hadamard 矩阵:

$$H_n \equiv H_2 \otimes H_{n/2} = \begin{bmatrix} H_{n/2} & H_{n/2} \\ H_{n/2} & -H_{n/2} \end{bmatrix},$$

$$n = 2^k, \quad k = 2,3,\cdots \tag{17.6}$$

例 规范化对称 Hadamard 矩阵H_4:

$$H_4 \equiv H_2 \otimes H_2 = \begin{bmatrix} H_2 & H_2 \\ H_2 & -H_2 \end{bmatrix}$$

$$= \begin{bmatrix} 1 & 1 & 1 & 1 \\ 1 & -1 & 1 & -1 \\ 1 & 1 & -1 & -1 \\ 1 & -1 & -1 & 1 \end{bmatrix}. \tag{17.7}$$

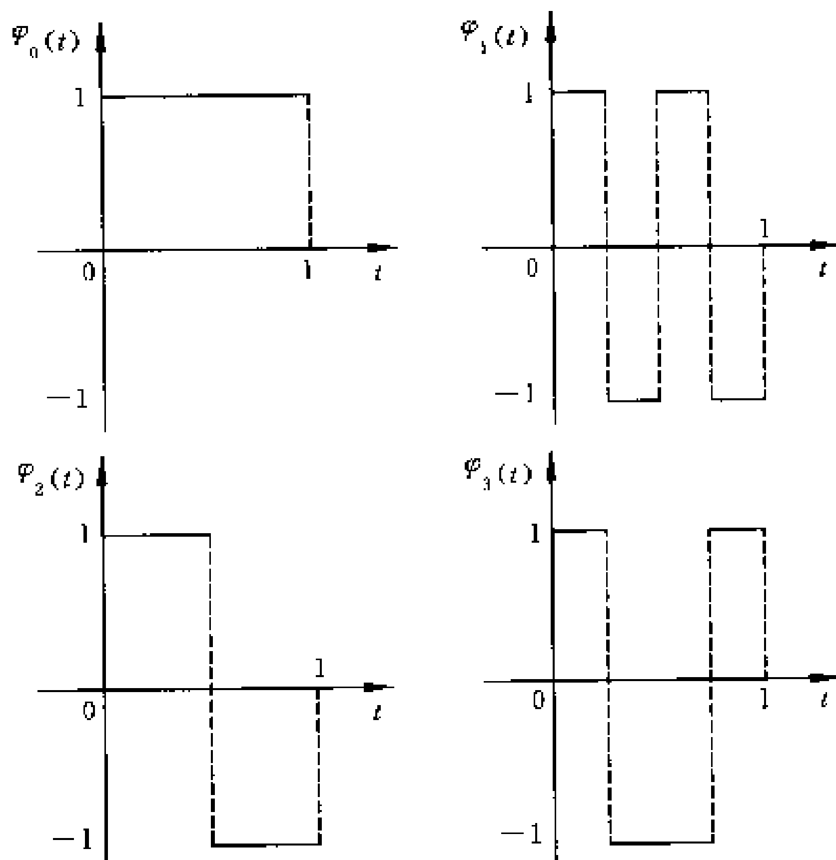
将区间$(0,1)$分作4等分,然后视H_4中每一行为区间$(0,1)$上的一个逐段线性函数,并依此记之为

$$\varphi_0, \varphi_1, \varphi_2, \varphi_3 .$$

例如,H_4中第2行对应的函数φ_1的表达式是

$$\varphi_1(t)=\begin{cases}1, & 0<t<1/4,\\ -1, & 1/4<t<1/2,\\ 1, & 1/2<t<3/4,\\ -1, & 3/4<t<1.\end{cases}$$

函数$\varphi_0,\varphi_1,\varphi_2,\varphi_3$的图象如下:



容易看出，$\varphi_0,\varphi_1,\varphi_2,\varphi_3$这四个矩形脉冲函数是正交的,即满足

$$\int_0^1 \varphi_i(t)\varphi_j(t)dt = \begin{cases} 0, & i \neq j, \\ 1, & i = j. \end{cases} \tag{17.8}$$

它们是在信息论和信号处理中有许多应用的 **Walsh 函数**的例子.

10.17.11 定义 设 $A \in \mathbb{C}^{m\times m}, B \in \mathbb{C}^{n\times n}$,而且 $b^{(1)},\cdots,b^{(n)}$ 是 B 的行.用 $A \overline{\otimes} B$ 表示依此以

$$A \otimes b^{(1)},\cdots, A \otimes b^{(n)}$$

为行而组成的 $mn\times mn$ 矩阵,块形式如下:

$$A \overline{\otimes} B \equiv \begin{bmatrix} A \otimes b^{(1)} \\ \vdots \\ A \otimes b^{(n)} \end{bmatrix},$$

其中 $\otimes$ 是通常 Kronecker 积的记号. $A \overline{\otimes} B$ 称为矩阵 A 和 B 的 **Paley 积**.

10.17.12 定理 设 A 和 B 是 Hadamard 矩阵,则 Paley 积 $A \overline{\otimes} B$ 也是 Hadamard 矩阵.

证明留作练习.

关于 Hadamard 矩阵的更深入的内容,可参看[A1975](包括应用于组合学),[HS1979](包括应用于光谱学和图象处理).

10.17.13 定义 设 $X_1,\cdots,X_n$ 是某概率空间上具有期望算子 E 的实或复随机变量,它们有有限第二矩,并用 $\mu_i \equiv E(X_i)$ 表示各自的平均值.自伴矩阵

$$\operatorname{cov}(X) \equiv \left[E\left((X_i - \mu_i)\left(\overline{X_j} - \overline{\mu_j}\right)\right)\right] \tag{17.9}$$

称为随机(变量)向量 $X = (X_1,\cdots,X_n)^{\mathrm{T}}$ 的**协方差矩阵**(covariance matrix).

注意,(17.9)中 $E\left(\left(X_i-\mu_i\right)\left(\overline{X_j}-\overline{\mu_j}\right)\right)$是矩阵 $\operatorname{cov}(X)$ 的 (i,j)-元素.

10.17.14 定理 设 $S=[s_{ij}]\in\mathbb{C}^{n\times n}$, $X=\left(X_1,\cdots,X_n\right)^{\mathrm{T}}$ 是随机向量,则

$$\operatorname{cov}(SX)=S\operatorname{cov}(X)S^*. \tag{17.10}$$

证 SX 也是随机向量,其分量是 X 的分量的线性组合.注意到期望算子 E 是线性的, SX 的分量的平均值是

$$E\left((SX)_i\right)=E\left(\sum_{k=1}^{n}s_{ik}X_k\right)=\sum_{k=1}^{n}s_{ik}E\left(X_k\right)=\sum_{k=1}^{n}s_{ik}\mu_k .$$

从而 SX 的协方差矩阵是

$$\begin{aligned}\operatorname{cov}(SX)&=\left[E\left(\left((SX)_i-E\left((SX)_i\right)\right)\left(\overline{(SX)_j}-E\left(\overline{(SX)_j}\right)\right)\right)\right]\\&=\left[E\left(\left(\sum_{p=1}^{n}s_{ip}\left(X_p-\mu_p\right)\right)\left(\sum_{q=1}^{n}\overline{s_{jq}}\left(\overline{X_q}-\overline{\mu_q}\right)\right)\right)\right]\\&=\left[\sum_{p,q=1}^{n}s_{ip}E\left(\left(X_p-\mu_p\right)\left(\overline{X_q}-\overline{\mu_q}\right)\right)\overline{s_{jq}}\right]\\&=S\operatorname{cov}(X)S^* .\end{aligned}$$

□

10.17.15 定理 设 $X=\left(X_1,\cdots,X_n\right)^{\mathrm{T}}$ 和 $Y=\left(Y_1,\cdots,Y_n\right)^{\mathrm{T}}$ 是实或复四阶向量随机变量,而且具有零平均值,即有

$$E(X)\equiv\left(E\left(X_1\right),\cdots,E\left(X_n\right)\right)^{\mathrm{T}}=0$$

和

$$E(Y)\equiv\left(E\left(Y_1\right),\cdots,E\left(Y_n\right)\right)^{\mathrm{T}}=0,$$

这里 E 是期望算子,则

(1) X 的协方差矩阵

$$\operatorname{cov}(X)=\left[E\left(X_i\overline{X_j}\right)\right],$$

而且是半正定的.

(2) 如果 X 和 Y 是独立的,那么作为 X 和 Y 的 Hadamard 积二阶向量随机变量 $Z=X\circ Y$ 具有零平均值,而且 Z 的协方差矩阵等于 X 的协方差矩阵和 Y 的协方差矩阵的 Hadamard 积,即

$$\operatorname{cov}(Z)=\operatorname{cov}(X)\circ\operatorname{cov}(Y).$$

证 (1) 根据(17.9),

$$\begin{aligned}\operatorname{cov}(X)&=\left[E\left(\left(X_i-\mu_i\right)\left(\overline{X_j}-\overline{\mu_j}\right)\right)\right]\\&\equiv E\left[\left(X-E(X)\right)\left(X-E(X)\right)^*\right].\end{aligned}$$

由此,因 $E(X)=0$,得

$$\operatorname{cov}(X)=E\left(XX^*\right)=\left[E\left(X_i\overline{X_j}\right)\right].$$

而且,鉴于期望算子具有线性、齐次性和非负性,对任何

$$\xi=\left(\xi_1,\cdots,\xi_n\right)^{\mathrm{T}}\in\mathbb{C}^n,$$

有

$$\begin{aligned}\xi^*\operatorname{cov}(X)\xi&=\sum_{i,j=1}^n E\left(X_i\overline{X_j}\right)\overline{\xi_i}\xi_j=\sum_{i,j=1}^n E\left(\overline{\xi_i}X_i\xi_j\overline{X_j}\right)\\&=E\left(\sum_{i,j=1}^n\overline{\xi_i}X_i\xi_j\overline{X_j}\right)=E\left(\left|\sum_{i=1}^n\overline{\xi_i}X_i\right|^2\right)\ge 0,\end{aligned}$$

故 $\operatorname{cov}(\ )$ 是半正定的.

(2) 由 $E(X)=0$, $E(Y)=0$,及 X 和 Y 是独立的,直接推出

$$\begin{aligned}E(Z)&=E(X\circ Y)=\left(E\left(X_1Y_1\right),\cdots,E\left(X_nY_n\right)\right)^{\mathrm{T}}\\&=\left(E\left(X_1\right)E\left(Y_1\right),\cdots,E\left(X_n\right)E\left(Y_n\right)\right)^{\mathrm{T}}=0.\end{aligned}$$

于是,根据(1),

$$\operatorname{cov}(Z)=E\left(ZZ^*\right)=\left[E\left(X_iY_i\overline{X_j}\overline{Y_j}\right)\right]=\left[E\left(X_i\overline{X_j}Y_i\overline{Y_j}\right)\right]$$

$$=\left[E\left(X_i\overline{X_j}\right)E\left(Y_i\overline{Y_j}\right)\right]=\operatorname{cov}(X)\circ\operatorname{cov}(Y). \qquad \square$$

10.17.16 定义 $A=[a_{ij}]\in\mathbb{C}^{n\times n}$ 称为**衡平矩阵**(equitable matrix),如果满足

$$a_{ij}a_{jk}=a_{ik}, \quad \forall i,j,k=1,\cdots,n. \tag{17.11}$$

此类矩阵在经济学及群论中有着应用.

10.17.17 定理 设 $A=[a_{ij}]\in\mathbb{C}^{n\times n}$ 是衡平矩阵,则

(1) A 形如

$$A=\begin{pmatrix}1\\a_{21}\\a_{31}\\\vdots\\a_{n1}\end{pmatrix}\left(1,\frac{1}{a_{21}},\frac{1}{a_{31}},\cdots,\frac{1}{a_{n1}}\right)$$

$$=\begin{bmatrix}1&\frac{1}{a_{21}}&\frac{1}{a_{31}}&\cdots&\frac{1}{a_{n1}}\\a_{21}&1&\frac{a_{21}}{a_{31}}&\cdots&\frac{a_{21}}{a_{n1}}\\a_{31}&\frac{a_{31}}{a_{21}}&1&\cdots&\frac{a_{31}}{a_{n1}}\\\cdots&\cdots&\cdots&\cdots&\cdots\\a_{n1}&\frac{a_{n1}}{a_{21}}&\frac{a_{n1}}{a_{31}}&\cdots&1\end{bmatrix}. \tag{17.12}$$

(2) $A^2=nA$.

(3) 成立

$$R^{-1}AR=(n,0,\cdots,0), \tag{17.13}$$

其中

$$R=\begin{bmatrix}1 & 0 & \cdots & \cdots & 0 & -1\\ a_{21} & a_{21} & 0 & & & 0\\ a_{31} & -a_{31} & a_{31} & \ddots & & \vdots\\ a_{41} & 0 & -a_{41} & \ddots & \ddots & \vdots\\ \vdots & \vdots & \ddots & \ddots & \ddots & 0\\ a_{n1} & 0 & \cdots & 0 & -a_{n1} & a_{n1}\end{bmatrix}. \qquad (17.14)$$

证 根据(17.11),

$$a_{ik}=a_{ii}a_{ik}, \quad k=1,\cdots,n, \ i=1,\cdots,n,$$

得

$$a_{ii}=1, \quad i=1,\cdots,n.$$

又$1=a_{ii}=a_{i1}a_{1i}, i=1,\cdots,n$,得

$$a_{1i}=\frac{1}{a_{i1}}, \quad i=1,\cdots,n.$$

因此

$$a_{ik}=a_{i1}a_{1k}=\frac{a_{i1}}{a_{k1}}, \quad i=1,\cdots,n.$$

从而(17.12)成立.

利用(17.12),

$$A^2=\begin{pmatrix}1\\ a_{21}\\ \vdots\\ a_{n1}\end{pmatrix}\left(1,\frac{1}{a_{21}},\cdots,\frac{1}{a_{n1}}\right)\begin{pmatrix}1\\ a_{21}\\ \vdots\\ a_{n1}\end{pmatrix}\left(1,\frac{1}{a_{21}},\cdots,\frac{1}{a_{n1}}\right)$$

$$=\begin{pmatrix}1\\ a_{21}\\ \vdots\\ a_{n1}\end{pmatrix}\left\{\left(1,\frac{1}{a_{21}},\cdots,\frac{1}{a_{n1}}\right)\begin{pmatrix}1\\ a_{21}\\ \vdots\\ a_{n1}\end{pmatrix}\right\}\left(1,\frac{1}{a_{21}},\cdots,\frac{1}{a_{n1}}\right)$$

$$
=\left\{\left(1,\frac{1}{a_{21}},\cdots,\frac{1}{a_{n1}}\right)\begin{pmatrix}1\\a_{21}\\\vdots\\a_{n1}\end{pmatrix}\right\}\begin{pmatrix}1\\a_{21}\\\vdots\\a_{n1}\end{pmatrix}\left(1,\frac{1}{a_{21}},\cdots,\frac{1}{a_{n1}}\right)
$$

$$
=nA.
$$

(2)得证.现在证明(3).由于

$$
R^{-1}AR=R^{-1}\begin{pmatrix}1\\a_{21}\\\vdots\\a_{n1}\end{pmatrix}\left(1,\frac{1}{a_{21}},\cdots,\frac{1}{a_{n1}}\right)R,
$$

容易算得

$$
\left(1,\frac{1}{a_{21}},\cdots,\frac{1}{a_{n1}}\right)R=(n,0,\cdots,0).
$$

另一方面,令

$$
\begin{pmatrix}x_1\\x_2\\\vdots\\x_n\end{pmatrix}\equiv R^{-1}\begin{pmatrix}1\\a_{21}\\\vdots\\a_{n1}\end{pmatrix},
$$

等价于方程组

$$
\begin{cases}
x_1 & & & & & = 1,\\
x_1 & +x_2 & & & & = 1,\\
x_1 & -x_2 & +x_3 & & & = 1,\\
x_1 & & -x_3 & +x_4 & & = 1,\\
\cdots & & & & & \\
x_1 & & & -x_{n-1} & +x_n & = 1,
\end{cases}
$$

其解为

$$(x_1,x_2,\cdots,x_n)^{\mathrm{T}}=(1,0,\cdots,0)^{\mathrm{T}}.$$

于是

$$R^{-1}AR=(1,0,\cdots,0)^{\mathrm{T}}(n,0,\cdots,0)=\mathrm{diag}(n,0,\cdots,0).\qquad\square$$

例 $n\times n$矩阵

$$E=\begin{pmatrix}1\\1\\\vdots\\1\end{pmatrix}(1,1,\cdots,1)=\begin{bmatrix}1&1&\cdots&1\\1&1&\cdots&1\\\cdots&\cdots&\cdots&\cdots\\1&1&\cdots&1\end{bmatrix}$$

是衡平矩阵.

证明留作练习.

10.17.18 定义 在量子力学中出现的矩阵

$$S_1\equiv\begin{bmatrix}0&1\\1&0\end{bmatrix},\quad S_2\equiv\begin{bmatrix}0&-i\\i&0\end{bmatrix},\quad S_3\equiv\begin{bmatrix}1&0\\0&-1\end{bmatrix},$$

称为 **Pauli 旋量矩阵**(Pauli spin matrix).

10.17.19 定理 设S_1,S_2,S_3是 **10.17.18** 定义的旋量矩阵,则

(1) $\forall a_1,a_2,a_3\in\mathbb{R}$,成立

$$(a_1S_1+a_2S_2+a_3S_3)^2=a^2I$$

和

$$\mathrm{e}^{a_1S_1+a_2S_2+a_3S_3}=(\cosh a)I+\frac{\sinh a}{a}(a_1S_1+a_2S_2+a_3S_3),$$

其中$a\equiv(a_1^2+a_2^2+a_3^2)^{1/2}$.

(2) 如果$A\equiv tS_1,B\equiv tS_2,C\equiv t(S_3-S_2-S_1)$,那么成立

$$\mathrm{tr}(\mathrm{e}^{A+B+C})=2\cosh t$$

和

$$\operatorname{tr}\left(e^{A}e^{B}e^{C}\right)=2\left(1-\frac{t^{4}}{12}+O\left(t^{6}\right)\right)\cosh t.$$

证明留作练习.

10.17.20 定义 对于不是纯虚数或零的 $x\in\mathbb{C}$,定义**正负号函数**(sign function)为

$$\operatorname{sgn}x\equiv\begin{cases}+1, & \operatorname{Re}x>0,\\ -1, & \operatorname{Re}x<0.\end{cases}\tag{17.15}$$

这是按矩阵的说法来定义的一种纯量的正负号函数.

10.17.21 定义 设 $A\in\mathbb{C}^{n\times n}$ 没有纯虚数或零的特征值.并设 J 是 A 的 Jordan 标准形,

$$A=PJP^{-1},\tag{17.16}$$

其中 $P\in\mathbb{C}^{n\times n}$ 是非奇异矩阵.

定义**矩阵正负号函数**(matrix sign function)为

$$\operatorname{sgn}A\equiv P(\operatorname{sgn}J)P^{-1},\tag{17.17}$$

其中

$$\begin{aligned}\operatorname{sgn}J\equiv\operatorname{diag}(\operatorname{sgn}\lambda_1,\cdots,\operatorname{sgn}\lambda_n),\\ \lambda_i\in\lambda(A),\quad i=1,\cdots,n.\end{aligned}\tag{17.18}$$

例 矩阵

$$A=\begin{bmatrix}1 & 3\\ 2 & 2\end{bmatrix}$$

的特征值为

$$\lambda_1=-1,\lambda_2=4.$$

对应的特征向量可取为

$$x^{(1)} = \begin{pmatrix} 3 \\ -2 \end{pmatrix}, \quad x^{(2)} = \begin{pmatrix} 1 \\ 1 \end{pmatrix}.$$

从而 A 的 Jordan 标准形为 $J = \operatorname{diag}(-1,4)$,

$$A = P\operatorname{diag}(-1,4)P^{-1},$$

其中

$$P = \begin{bmatrix} 3 & 1 \\ -2 & 1 \end{bmatrix}, \quad P^{-1} = \frac{1}{5}\begin{bmatrix} 1 & -1 \\ 2 & 3 \end{bmatrix}.$$

根据(17.18),

$$\operatorname{sgn} J = \operatorname{diag}(-1,1).$$

然后根据(17.17),

$$\operatorname{sgn} A = P\begin{bmatrix} -1 & 0 \\ 0 & 1 \end{bmatrix}P^{-1} = \frac{1}{5}\begin{bmatrix} -1 & 6 \\ 4 & 1 \end{bmatrix}.$$

10.17.22 定理 设 $A \in \mathbb{C}^{n\times n}$,矩阵正负号函数有下列基本性质:

(1) 若 A 的所有特征值全具有正实部或全具有负实部,则分别地 $\operatorname{sgn} A = I$ 或 $\operatorname{sgn} A = -I$.

(2) 若 A 的 Jordan 标准形是对角矩阵,而且所有特征值是 ± 1,则 $\operatorname{sgn} A = A$.

(3) $(\operatorname{sgn} A)^2 = I$.

证明留作练习.

从定义 **10.17.21** 出发计算 $\operatorname{sgn} A$ 是不现实的,因为要求知道 A 的特征值的实部以及变换矩阵 P.一个比较好的计算方案是以非线性方程求根的 Newton-Raphson 方法为基础的,可参看[B1990].

10.17.23 定义 设矩阵 $A\in[a_{ij}]\in\mathbb{R}^{n\times n}$,矩阵 $B=[b_{ij}]\in\mathbb{R}^{n\times n}$ 关于 A 称为**正负号相似**的(sign similar),如果

$$\begin{cases} \operatorname{sgn} b_{ij}=\operatorname{sgn} a_{ij}, & \text{当 } a_{ij}\neq 0, \\ b_{ij}=0, & \text{当 } a_{ij}=0. \end{cases}$$

在一些经济模型中出现的矩阵常常仅知道元素(不包括零)的正负号.

10.17.24 定义 设矩阵 $A\in\mathbb{R}^{n\times n}$.如果存在稳定矩阵 $B\in\mathbb{R}^{n\times n}$ 正负号相似于 A,则称 A 为**潜在稳定矩阵**(potentially stable matrix).

10.17.25 定义 设矩阵 $A\in\mathbb{R}^{n\times n}$.如果正负号相似于 A 的每个矩阵 $B\in\mathbb{R}^{n\times n}$ 都是稳定矩阵,则称 A 为**正负号稳定矩阵**(sign stable matrix).

显然,对于正负号稳定矩阵 A,不论怎样改变其元素的值,只要保持其正、负和零元素的分布不变,都必定是稳定矩阵(自然包括 A 本身).

关于正负号相似、潜在稳定矩阵和正负号稳定矩阵的内容可参看[B1984].

10.17.26 定义 设 $A(t)=\left[a_{ij}(t)\right]$ 是 $t\in\left[0,t_0\right]$ 的连续的 $n\times n$ 函数矩阵.

微分方程一般理论保证如下一阶齐次线性微分方程组初值问题

$$\begin{cases} \dfrac{\mathrm{d}}{\mathrm{d}t}X(t)=A(t)X(t), & t\in\left(0,t_0\right], \\ X(0)=I, \end{cases} \tag{17.19}$$

有唯一解——$n\times n$ 函数矩阵——$X(t)$.

函数矩阵

$$G(t,s) \equiv X(t)\left(X(s)\right)^{-1} \tag{17.20}$$

称为初值问题(17.19)的 **Green 矩阵**.

10.17.27 定理 设 $X(t)$ 是(17.19)的解,且 $G(t,s)$ 是相应的 Green 矩阵,则

(1) 成立

$$\det X(t) = \exp\left(\int_0^t A(s)\mathrm{d}s\right), \quad \forall t \in [0,t_0).$$

(2) 对于所有 $t \in [0,t_0)$, $X(t)$ 是非奇异的.

(3) $G(t,t) = I$.

(4) 成立

$$\frac{\mathrm{d}}{\mathrm{d}t}G(t,s) = A(t)G(t,s).$$

(5) 如果 $y(t)$ 是已知的 $[0,t_0)$ 上的 n 维函数向量,那么 n 维函数向量

$$x(t) \equiv X(t)x(0) + \int_0^t G(t,s)y(s)\mathrm{d}s$$

是一阶非齐次线性问题

$$\begin{cases} \dfrac{\mathrm{d}}{\mathrm{d}t}x(t) = A(t)x(t) + y(t), & t \in (0,t_0), \\ x(t)\big|_{t=0} = x(0) \end{cases}$$

的解.

(6) 如果 $A \in \mathbb{C}^{n\times n}$ 是常数矩阵,那么

$$G(t,s) = \mathrm{e}^{(t-s)A}.$$

而且一阶非齐次线性问题

$$\begin{cases} \dfrac{\mathrm{d}}{\mathrm{d}t}x(t) = Ax(t) + y(t), & t \in (0,t_0), \\ x(t)\big|_{t=0} = x(0) \end{cases}$$

的解是

$$x(t)=\mathrm{e}^{A}x(0)+\mathrm{e}^{tA}\int_0^t \mathrm{e}^{-sA}y(s)\mathrm{d}s.$$

证明留作练习.

10.18 自反矩阵

10.18.1 定义 $R\in\mathbb{C}^{n\times n}$ 称为**广义反射矩阵**(generalized reflection matrix),如果它满足如下两个条件:

(1) $R=R^*$;

(2) $R^2=I$.

换句话说,广义反射矩阵就是自伴的对合矩阵.

10.18.2 定义 设 $P\in\mathbb{C}^{n\times n}$ 是广义反射矩阵.

$x\in\mathbb{C}^n$ 分别地称为关于 P 的**自反向量**(reflexive vector)和**反自反向量**(antireflexive vector),如果分别地有

$$x=Px \quad 和 \quad x=-Px.$$

$A\in\mathbb{C}^{n\times n}$ 分别地称为关于 P 的**自反矩阵**(reflexive matrix) matrix)和**反自反矩阵**(antireflexive matrix),如果分别地有

$$A=PAP \quad 和 \quad A=-PAP.$$

$S\subset\mathbb{C}^n$ 分别地称为关于 P 的**自反子空间**(reflexive subspace) 和**反自反子空间**(antireflexive subspace),如果分别地有

$$x=Px,\forall x\in S \quad 和 \quad x=-Px,\forall x\in S.$$

并引进记号

$$\mathbb{C}_r^n(P)\equiv\{x:x=Px,x\in\mathbb{C}^n\} \tag{18.1}$$

和

$$\mathbb{C}_a^n(P) \equiv \{ x : x = -Px, x \in \mathbb{C}^n \}. \tag{18.2}$$

$S \subset \mathbb{C}^{n\times n}$ 分别地称为关于 P 的**自反子空间**和**反自反子空间**,如果分别地有

$$A = PAP, \forall A \in S \quad 和 \quad A = -PAP, \forall A \in S.$$

并引进记号

$$\mathbb{C}_r^{n\times n}(P) \equiv \{ A : A = PAP, A \in \mathbb{C}^{n\times n} \} \tag{18.3}$$

和

$$\mathbb{C}_a^{n\times n}(P) \equiv \{ A : A = -PAP, A \in \mathbb{C}^{n\times n} \}. \tag{18.4}$$

10.18.3 定义 设 $P \in \mathbb{C}^{m\times m}$ 和 $Q \in \mathbb{C}^{n\times n}$ 是广义反射矩阵.

$A \in \mathbb{C}^{m\times n}$ 称为关于矩阵对 (P,Q) 的**广义自反矩阵**(generalized reflexive matrix),如果

$$A = PAQ.$$

$A \in \mathbb{C}^{m\times n}$ 称为关于矩阵对 (P,Q) 的**广义反自反矩阵**(generalized antireflexive matrix),如果

$$A = -PAQ.$$

$S \subset \mathbb{C}^{m\times n}$ 称为关于矩阵对 (P,Q) 的**广义自反子空间**(generalized reflexive subspace),如果

$$A = PAQ, \ \forall A \in S.$$

$S \subset \mathbb{C}^{m\times n}$ 称为关于矩阵对 (P,Q) 的**广义反自反子空间**(generalized antireflexive subspace),如果

$$A = -PAQ, \ \forall A \in S.$$

引进记号

$$\mathbb{C}_r^{m\times n}(P,Q) \equiv \{ A : A = PAQ, A \in \mathbb{C}^{m\times n} \} \tag{18.5}$$

和

$$\mathbb{C}_a^{m\times n}(P,Q)\equiv\{A:A=-PAQ,A\in\mathbb{C}^{m\times n}\}. \qquad (18.6)$$

特别,

$$\mathbb{C}_r^{n\times n}(P)\equiv\mathbb{C}_r^{n\times n}(P,P)$$

和

$$\mathbb{C}_a^{n\times n}(P)\equiv\mathbb{C}_a^{n\times n}(P,P).$$

$A\in\mathbb{C}^{m\times n}$ 分别地称为具有**广义 SAS (generalized SAS)性质**和**广义反-SAS (generalized anti-SAS)性质**,如果 A 分别地是广义自反矩阵和广义反自反矩阵,其中 SAS 代表对称和反对称.

依如此定义,显然,自反(反自反)矩阵或向量必定是广义自反(广义反自反)矩阵.但是,反过来说,一般是不成立的.

值得指出的是,中心对称(斜中心对称)矩阵是广义自反(广义反自反)矩阵的特殊情形,而且通过广义自反(广义反自反)概念可以将中心对称(斜中心对称)定义从方阵推广至长方阵.

例 设

$$A=\begin{bmatrix}\alpha & \beta & \gamma\\ \mu & \nu & \nu\\ \alpha & \gamma & \beta\end{bmatrix},\quad P=\begin{bmatrix}0&0&1\\0&1&0\\1&0&0\end{bmatrix},\quad Q=\begin{bmatrix}1&0&0\\0&0&1\\0&1&0\end{bmatrix},$$

容易验证 $A=PAQ$.因此实或复矩阵 A 是关于 (P,Q) 的广义自反矩阵.

10.18.4 定义 矩阵 $X\in\mathbb{C}^{m\times n}$ 称为**左(或右)正交**于矩阵 $Y\in\mathbb{C}^{m\times n}$,如果

$$X^*Y=0\,(\text{或}\,YX^*=0).$$

子空间 $S_1\subset\mathbb{C}^{m\times n}$ 称为**左(或右)正交**于子空间 $S_2\subset\mathbb{C}^{m\times n}$,如果

$$X^*Y = 0\,(\text{或}\,YX^* = 0),\ \ \forall X \in S_1, Y \in S_2.$$

10.18.5 定理 设 $P \in \mathbb{C}^{m\times m}$ 和 $Q \in \mathbb{C}^{n\times n}$ 是广义反射矩阵,而且 $\alpha, \beta \in \mathbb{C}$.

(1) 若 $A, B \in \mathbb{C}_r^{m\times n}(P,Q)$,则

$$\alpha A^+ + \beta B^+ \in \mathbb{C}_r^{n\times m}(Q,P),$$

$$\alpha A^* + \beta B^* \in \mathbb{C}_r^{n\times m}(Q,P),$$

$$A^*B \in \mathbb{C}_r^{n\times n}(Q),\ \ AB^* \in \mathbb{C}_r^{m\times m}(P).$$

(2) 若 $A, B \in \mathbb{C}_a^{m\times n}(P,Q)$,则

$$\alpha A^+ + \beta B^+ \in \mathbb{C}_a^{n\times m}(Q,P),$$

$$\alpha A^* + \beta B^* \in \mathbb{C}_a^{n\times m}(Q,P),$$

$$A^*B \in \mathbb{C}_r^{n\times n}(Q),\ \ AB^* \in \mathbb{C}_r^{m\times m}(P).$$

(3) 若 $A \in \mathbb{C}_r^{m\times n}(P,Q)$和 $B \in \mathbb{C}_a^{m\times n}(P,Q)$,则

$$\alpha A^*A + \beta B^*B \in \mathbb{C}_r^{n\times n}(Q),$$

$$\alpha AA^* + \beta BB^* \in \mathbb{C}_r^{m\times m}(P),$$

$$A^*B \in \mathbb{C}_a^{n\times n}(Q),\ \ AB^* \in \mathbb{C}_a^{m\times m}(P).$$

这里 A^+ 和 B^+ 是 Moore-Penrose 逆.

证 依 Moore-Penrose 逆的定义 **7.6.2**,并利用 $A = PAQ$,得

$$PAQA^+PAQ = PAQ.$$

由此并注意到 P 和 Q 非奇异,有

$$AYA = A, \qquad Y \equiv QA^+P.$$

这表明 Y 满足 **7.6.2** 中 Moore-Penrose 四个条件的第一个条件.利用 P 和 Q 是酉自伴矩阵,容易证明 Y 也满足其余三个 Moore-Penrose 条件.因此,Y 是 A 的广义逆.因为 Moore-Penrose 逆是唯一的,所以

$$A^+ = Y \equiv QA^+P \in \mathbb{C}_r^{n\times m}(Q,P),$$

同样地,

$$B^{+} = Y \equiv QB^{+}P \in \mathbb{C}_{r}^{n\times m}(Q,P).$$

于是$\alpha A^{+} + \beta B^{+} \in \mathbb{C}_{r}^{n\times m}(Q,P)$.(1)的余下部分的证明没有新的困难,因而从略.

(2)和(3)的证明相仿,留作练习. □

10.18.6 定理 设$P \in \mathbb{C}^{m\times m}$和$Q \in \mathbb{C}^{n\times n}$是广义反射矩阵,则任一矩阵$A \in \mathbb{C}^{m\times n}$可以分解成两个部分$U$和$V$,

$$U + V = A,$$

使得

$$U \in \mathbb{C}_{r}^{m\times n}(P,Q), \quad V \in \mathbb{C}_{a}^{m\times n}(P,Q).$$

证 取

$$U \equiv \frac{1}{2}(A + PAQ), \quad V \equiv \frac{1}{2}(A - PAQ), \qquad (18.7)$$

显然$U + V = A$.然后,利用对合性质$P^{2} = I$和$Q^{2} = I$,得

$$PUQ = \frac{1}{2}P(A + PAQ)Q = \frac{1}{2}(PAQ + A) = U$$

和

$$PVQ = \frac{1}{2}P(A - PAQ)Q = \frac{1}{2}(PAQ - A) = -V. \qquad □$$

10.18.7 定理 $\mathbb{C}_{r}^{m\times n}(P,Q)$和$\mathbb{C}_{a}^{m\times n}(P,Q)$分别是域$\mathbb{C}$上$\mathbb{C}^{m\times n}$关于$(P,Q)$的广义自反子空间和广义反自反子空间.

而且,$\mathbb{C}_{r}^{m\times n}(P,I_{n})$左正交于$\mathbb{C}_{a}^{m\times n}(P,I_{n})$,$\mathbb{C}_{r}^{m\times n}(I_{m},Q)$右正交于$\mathbb{C}_{a}^{m\times n}(I_{m},Q)$.

证 一方面,从 **10.18.6** 得知$\mathbb{C}_{r}^{m\times n}(P,Q)$是$\mathbb{C}^{m\times n}$的非空子集.

另一方面,对于任意的$X,Y \in \mathbb{C}_{r}^{m\times n}(P,Q)$和任意的$\alpha,\beta \in \mathbb{C}$,因为

$$P(\alpha X+\beta Y)Q=\alpha PXQ+\beta PYQ=\alpha X+\beta Y,$$

表明

$$\alpha X+\beta Y\in\mathbb{C}_r^{m\times n}(P,Q),$$

所以$\mathbb{C}_r^{m\times n}(P,Q)$是域$\mathbb{C}$上$\mathbb{C}^{m\times n}$的子空间,而且根据(18.5),还是关于$(P,Q)$广义自反的.

类似地可证$\mathbb{C}_a^{m\times n}(P,Q)$是域$\mathbb{C}$上$\mathbb{C}^{m\times n}$关于$(P,Q)$的广义反自反子空间.

其次,对于任意的$X\in\mathbb{C}_r^{m\times n}(P,I_n)$和$Y\in\mathbb{C}_a^{m\times n}(P,I_n)$,有

$$X^*Y=(X^*P^*)(-PY)=-X^*Y,$$

推出$X^*Y=0$,从而

$$Y^*X=(X^*Y)^*=0.$$

因此,$\mathbb{C}_r^{m\times n}(P,I_n)$和$\mathbb{C}_a^{m\times n}(P,I_n)$相互左正交.

同样,$\mathbb{C}_r^{m\times n}(I_m,Q)$和$\mathbb{C}_a^{m\times n}(I_m,Q)$相互右正交. □

10.18.8 推论 $\mathbb{C}_r^n(P)$和$\mathbb{C}_a^n(P)$分别是域$\mathbb{C}$上$\mathbb{C}^n$关于P的广义自反子空间和广义反自反子空间.而且,$\mathbb{C}_r^n(P)$和$\mathbb{C}_a^n(P)$相互左正交.

$\mathbb{C}_r^{n\times n}(P)$和$\mathbb{C}_a^{n\times n}(P)$分别是域$\mathbb{C}$上$\mathbb{C}^{n\times n}$关于$P$的广义自反子空间和广义反自反子空间.

证 结论是显然的,因为自反子空间和反自反子空间必然分别是广义自反子空间和广义反自反子空间. □

注意,虽然$\mathbb{C}_r^n(P)$和$\mathbb{C}_a^n(P)$相互左正交,但是一般它们并不相互右正交.

10.18.9 定理 设给定一个线性最小二乘问题

$$\min_x\|Ax-b\|_2,$$

其中

$$A\in\mathbb{C}^{m\times n},\ x\in\mathbb{C}^{n},\ b\in\mathbb{C}^{m},\ n\le m.$$

而且 A 有满列秩,即 $\operatorname{rank}A=n$;并设 $\tilde{x}$ 是问题的唯一解,其剩余为

$$\tilde{r}\equiv b-A\tilde{x},$$

则

(1) 如果 $A\in\mathbb{C}_r^{m\times n}(P,Q)$,那么当 $b\in\mathbb{C}_r^m(P)$ 时,

$$\tilde{x}\in\mathbb{C}_r^n(Q),\quad \tilde{r}\in\mathbb{C}_r^m(P).$$

当 $b\in\mathbb{C}_a^m(P)$ 时,

$$\tilde{x}\in\mathbb{C}_a^n(Q),\quad \tilde{r}\in\mathbb{C}_a^m(P).$$

(2) 如果 $A\in\mathbb{C}_a^{m\times n}(P,Q)$,那么当 $b\in\mathbb{C}_a^m(P)$ 时,

$$\tilde{x}\in\mathbb{C}_r^n(Q),\quad \tilde{r}\in\mathbb{C}_a^m(P).$$

当 $b\in\mathbb{C}_r^m(P)$ 时,

$$\tilde{x}\in\mathbb{C}_a^n(Q),\quad \tilde{r}\in\mathbb{C}_r^m(P).$$

证 如果 $A\in\mathbb{C}_r^{m\times n}(P,Q)$,那么

$$A=PAQ.$$

且依定义,P 和 Q 是广义反射矩阵,从而有

$$P=P^*=P^{-1},Q=Q^*=Q^{-1}.$$

因 $\operatorname{rank}A=n$,故 A 的广义逆 A^+ 可以表示为

$$\begin{aligned}A^+&=(A^*A)^{-1}A^*=(QA^*PPAQ)^{-1}QA^*P\\&=(QA^*AQ)^{-1}QA^*P=Q(A^*A)^{-1}QQA^*P\\&=QA^+P.\end{aligned}$$

由此,在 $b=Pb$ 的条件下,依 **7.6.10**,有

$$\tilde{x}=A^+b=QA^+Pb=QA^+b=Q\tilde{x}$$

和

$$\tilde{r}=b-A\tilde{x}=Pb-PAQQ\tilde{x}=P(b-A\tilde{x})=P\tilde{r}.$$

类似地,在 $b=-Pb$ 的条件下,有

$$\tilde{x} = A^+b = QA^+Pb = -QA^+b = -Q\tilde{x}$$

和

$$\tilde{r} = b - A\tilde{x} = -Pb - PAQ(-Q\tilde{x}) = -P(b - A\tilde{x}) = -P\tilde{r}.$$

至此,(1)得证.

(2)的证明是类似的,留作练习. □

注意,上述定理中结论(1)和(2)之逆一般是不成立的.例如,假设

$$b \in \mathbb{C}^m, \quad b \notin \mathbb{C}_r^m(P) \text{且} b \notin \mathbb{C}_a^m(P).$$

将 **10.18.6** 应用于向量,可把 b 分解成两个非零向量 b_1 和 b_2,使得

$$b = b_1 + b_2,$$

其中 $b_1 \in \mathbb{C}_r^m(P)$ 和 $b_2 \in \mathbb{C}_a^m(P)$.令

$$\tilde{x}_1 = A^+b_1, \quad \tilde{x}_2 = A^+b_2,$$

则有

$$\tilde{x} = A^+b = A^+(b_1 + b_2) = \tilde{x}_1 + \tilde{x}_2,$$

其中 $\tilde{x}_1 \in \mathbb{C}_r^n(Q), \tilde{x}_2 \in \mathbb{C}_a^n(Q)$.

现在,如果选取 b_2 属于 A^* 的零空间,那么

$$\tilde{x}_2 = A^+b_2 = (A^*A)^{-1}A^*b_2 = 0,$$

于是

$$\tilde{x} = \tilde{x}_1 \in \mathbb{C}_r^n(Q).$$

同样,如果选取 b_1 属于 A^* 的零空间,那么

$$\tilde{x} = \tilde{x}_2 \in \mathbb{C}_a^n(Q).$$

还应注意,若矩阵 A 不是满秩的,则非极小范数解的最小二乘解(不唯一)可能没有上述定理所说的性质.构造这样的例子并不困难,最简单的情形是让 A 的某一列元素全为零(留作练习).

在 P 和 Q 确定之后,对于系数矩阵是关于 (P,Q) 广义自反的线

性最小二乘问题,**10.18.6** 和 **10.18.9** 可以提供将其分解成两个独立较小子问题的完整的重要信息.

例 考虑超定线性方程组

$$Ax=b;\quad A=\begin{bmatrix}4&2\\1&3\\2&4\\3&1\end{bmatrix},x=\begin{pmatrix}x_1\\x_2\end{pmatrix},b=\begin{pmatrix}b_1\\b_2\\b_3\\b_4\end{pmatrix}\qquad(18.8)$$

的线性最小二乘解.讨论如下两种情形:

(1) $b=(16,15,16,15)^{\mathrm{T}}$.设

$$P=\begin{bmatrix}0&I_2\\I_2&0\end{bmatrix},\qquad Q=\begin{bmatrix}0&1\\1&0\end{bmatrix},\qquad(18.9)$$

容易验证

$$A=PAQ,\quad b=Pb.$$

从 **10.18.9** 得知 $x=Qx$,即有

$$x_1=x_2.$$

因此,求解(18.8)等价于求解

$$\begin{pmatrix}6\\4\end{pmatrix}(x_1)=\begin{pmatrix}16\\15\end{pmatrix}.\qquad(18.10)$$

现在,(18.10)的最小二乘解是 $x_1=3$.

于是,原始问题的解是

$$x_1=x_2=3.$$

这可以通过求解**正规方程**

$$A^{\mathrm{T}}Ax=A^{\mathrm{T}}b$$

来检验.剩余

$$r=b-Ax=(-2,3,-2,3)^{\mathrm{T}},$$

显然满足 $r=Pr$,故 r 是关于 P 自反的.

(2) $b=(19,14,13,16)^{\mathrm{T}}$.仍取(18.9)中的 P 和 Q.此时,b 既不是自反的,也不是反自反的.

为了利用 A 的广义自反性质,首先将 b 分解成 u (是自反的)和 v(是反自反的),亦即使得

$$u=Pu,\quad v=-Pv;\quad u+v=b$$

将 **10.18.6** 应用于向量 b,得

$$u=(16,15,16,15)^{\mathrm{T}},\quad v=(3,-1,-3,1)^{\mathrm{T}}.$$

这样,因为

$$x=A^{+}b=A^{+}(u+v)=A^{+}u+A^{+}v\equiv y+z,$$

从而直接把求解 $Ax=b$ 归结为求解

$$Ay=u\quad 和\quad Az=v.$$

然而,如果不去利用 A 的广义 SAS 性质,如此分解便没有什么好处,反而需花费双倍的计算工作量.

下一步是利用 **10.18.9** 约化 $Ay=u$ 和 $Az=v$ 的规模.

先看 $Ay=u$.由于它正好是情形(1)的 $Ax=b$,直接搬用(1)的结果,其解为 $y=(3,3)^{\mathrm{T}}$.

再看 $Az=v$.对于它的分解类似于情形(1)的 $Ax=b$,只不过现在需要使用的是 v 的反自反性.从 **10.18.9**,并因 $A=PAQ$ 和 $v=-Qv$,得知

$$z=-Pz.$$

由此,若令 $z=(z_1,z_2)^{\mathrm{T}}$,则有

$$z_1=-z_2.$$

于是,求解 $Az=v$ 约化为求解

$$\begin{pmatrix}2\\-2\end{pmatrix}(z_1)=\begin{pmatrix}-3\\1\end{pmatrix},$$

此方程组的最小二乘解是 $z_1 = 1$.如此,$\boldsymbol{A}\boldsymbol{z} = \boldsymbol{v}$ 的最小二乘解是 $z = (1,-1)^{\mathrm{T}}$.

最后,原始问题的最小二乘解是

$$x = y + z = \begin{bmatrix} 4 \\ 2 \end{bmatrix}.$$

仍可通过求解原始方程组的正规方程 $\boldsymbol{A}^{\mathrm{T}}\boldsymbol{A}\boldsymbol{x} = \boldsymbol{A}^{\mathrm{T}}\boldsymbol{b}$ 来检验.

[C1998]引入并讨论了广义白反矩阵,还介绍了在物理问题中的应用.

数 学 符 号

$\{x : p(x)\}$　使命题 $p(x)$ 成立的所有 x 构成的集合

$f : X \to Y$　f 是从集合 X 到集合 Y 的映射

$f : x \mapsto y$　f 是元素 x 对应元素 y 的映射

$f : x \leftrightarrow y$　映射 f 是元素 x 对应元素 y 的一一对应

$C^n[a,b]$　闭区间 $[a,b]$ 上具有连续 n 阶导数的函数全体的集合

$C^n(a,b)$　开区间 (a,b) 上具有连续 n 阶导数的函数全体的集合

P_{n-1}　次数小于 n 的多项式全体的集合　1.1.2//2

$\mathbb{R}$　实数全体的集合　1.1.2//2

$\mathbb{C}$　复数全体的集合　1.1.2//2

$\mathbb{R}^n$　具有 n 个实分量的列向量全体的集合　1.1.2//2

$\mathbb{C}^n$　具有 n 个复分量的列向量全体的集合　1.1.2//2

$\overline{\alpha}$　复数 $\alpha = a + \mathrm{i}b$ 的共轭复数, $\overline{\alpha} = a - \mathrm{i}b$

$\mathrm{Re}\,\alpha$　复数 $\alpha = a + \mathrm{i}b$ 的实部, $\mathrm{Re}\,\alpha = a$

$\mathrm{Im}\,\alpha$　复数 $\alpha = a + \mathrm{i}b$ 的虚部, $\mathrm{Im}\,\alpha = b$

$\mathrm{sgn}(\cdot)$　符号函数, $x > 0$ 时 $\mathrm{sgn}\,x = +1$; $x < 0$ 时 $\mathrm{sgn}\,x = -1$

$\min X$ 或 $\min\limits_{x \in X} x$　实数集合 X 的最(极)小值

$\max X$ 或 $\max\limits_{x \in X} x$　实数集合 X 的最(极)大值

$\begin{pmatrix} n \\ k \end{pmatrix}$　组合数或称二项式展开系数

$\forall$　"对所有的……",或"对每一个……"

$x \in X$　元素 x 属于集合 X

$\emptyset$　空集

$X \cup Y$　集合 X 与集合 Y 的交集

$X \cap Y$　集合 X 与集合 Y 的并集

$Y \subset X, X \supset Y$　集合 X 包含集合 Y

$p \Rightarrow q, q \Leftarrow p$　命题 p 蕴涵命题 q，

或者说 p 是 q 的充分条件,或者说 q 是 p 的必要条件

$p \Leftrightarrow q$　命题 p 和命题 q 等价,或者说 p 的充分必要条件是 q

□　定理证明结束号,也用于不证自明的情形

$x=(x_1,\cdots,x_n)^{\mathrm{T}}$　x 是具有 n 个分量 $x_1,\cdots,x_n$ 的列向量

$x=(x_1,\cdots,x_n)$　x 是具有 n 个分量 $x_1,\cdots,x_n$ 的行向量

$\mathrm{span}\{x^{(1)},\cdots,x^{(m)}\}$　向量 $x^{(1)},\cdots,x^{(m)}$ 的生成空间　1.1.8//4

$\dim X$　线性空间 X 的维数　1.1.13//5

$\oplus$　直和运算符　1.1.16//6, 2.1.4//117

mod　模　1.1.17//7

$\{x\}$　同余类　1.1.17//7

$\{x_k\},\{x^{(k)}\},\{X^{(k)}\}$　数列、向量序列、矩阵序列

X/Y 或 $X(\mathrm{mod}\,Y)$　商空间　1.1.18//7

X'　线性空间 X 的对偶空间　1.2.2//9

$(\bullet,\bullet)$　双线性函数　1.2.5//12,　1.7.2//73

内积　1.7.2//73, 1.7.28//86

$Y^{\perp}$　线性子空间 Y 的零化子　1.2.6//12

线性子空间 Y 的正交补　1.7.14//78

$\mathrm{codim}\,Y$　线性子空间 Y 的余维数　1.2.7//12

$\mathscr{R}_T$　线性映射 T 的值域　1.3.1//14

$\mathscr{N}_T$　线性映射 T 的零空间(核)　1.3.2//15

$\mathscr{L}(X,Y)$　$X \to Y$ 的线性映射空间　1.3.6//19

$S \circ T$　线性映射 T 和 S 的复合　1.3.6//19

$\circ$　Hadamard 乘号　7.2.1//441

T^{-1}　线性映射 T 的逆映射　1.3.7//20

矩阵 T 的逆矩阵　1.4.5//32

T'　线性映射T的转置　1.3.8//20

I　单位矩阵　1.4.6//33;　或恒等映射　1.3.11//22

O　零映射　1.3.11//23

$e_1,\cdots,e_n$　$\mathbb{C}^n$(或$\mathbb{R}^n$)中的标准基　1.4.1//26

$E\equiv\begin{bmatrix}e_n & \cdots & e_1\end{bmatrix}$　反序单位矩阵　2.1.11//120

$$T\equiv\begin{bmatrix}t_{ij}\end{bmatrix}\equiv\begin{bmatrix}t_{11} & \cdots & t_{1n}\\ \cdots & \cdots & \cdots\\ t_{m1} & \cdots & t_{mn}\end{bmatrix}\equiv\begin{bmatrix}c^{(1)} & \cdots & c^{(n)}\end{bmatrix}\equiv\begin{bmatrix}r^{(1)}\\ \vdots\\ r^{(m)}\end{bmatrix}$$

$m\times n$矩阵的几种表示　1.4.2//26-28

$(T)_{ij}$　矩阵T的(i,j)-元素t_{ij}　1.4.2//27

$T\in\mathbb{R}^{m\times n}$　T是$m\times n$实矩阵　1.4.2//27

$T\in\mathbb{C}^{m\times n}$　T是$m\times n$复矩阵　1.4.2//27

0　零矩阵、零向量、数零　1.4.2//28

T^{T}　矩阵T的转置　1.4.3//31

rank$T$　矩阵T的秩　1.4.4//31

diag$(d_1,\cdots,d_n)$　具有对角元素$d_1,\cdots,d_n$的对角矩阵　1.4.6//33

det$A$　$n\times n$矩阵A的行列式　1.5.7//40

A_{ij}　$n\times n$矩阵A删去第i行和第j列后的子矩阵　1.5.10//42

adj$A$　$n\times n$矩阵A的伴随矩阵　1.5.13//46

tr$A$　$n\times n$矩阵A的迹　1.5.15//47

$G_i(A)$　$n\times n$矩阵A的Gerschgorin圆盘　1.6.23//68

$G(A)$　$n\times n$矩阵A的Gerschgorin域　1.6.23//68

$\|\bullet\|$　范数　1.7.21//82, 1.8.2//89, 1.8.21//98,1.8.28//101, 1.8.30//102

$\|\bullet\|_1$　1-范数　1.8.8//90, 1.8.31//103

$\|\bullet\|_2$　2-范数　1.8.8//90, 1.8.31//103

$\|\bullet\|_\infty$　∞-范数　1.8.8//90, 1.8.31//103

$\|\cdot\|_p \ (p \geq 1)$　p-范数　1.8.8//91

$\|\cdot\|_F$　矩阵的 Frobenius 范数

P_Y　正交投影映射　1.7.16//79

A^*　矩阵(线性映射) A 的伴随矩阵(映射)　1.7.19//80

　　亦即矩阵 A 的共轭转置矩阵　1.7.28//87

$A(\alpha,\beta)$　矩阵 A 的子矩阵(α 和 β 是指标集)　2.1.1//113

$A(\alpha)$　矩阵 A 的主子矩阵(α 是指标集)　2.1.1//113

$A(\{1,\cdots,i\})$　矩阵 A 的主子矩阵　2.1.1//113

$\lambda(A)$　$n\times n$ 矩阵 A 的特征值全体的集合　2.1.19//123

$\lambda_i(A)$　$n\times n$ 矩阵 A 的第 i 个特征值

$\rho(A)$　$n\times n$ 矩阵 A 的谱半径　2.1.19//123

$n(\lambda_i)$　特征值 λ_i 的代数重数　2.1.29//128

$m(\lambda_i)$　特征值 λ_i 的几何重数　2.1.29//128

$\kappa(A)$　$n\times n$ 矩阵 A 的条件数　2.5.2//157

$V(x_1,\cdots,x_n)$　数组 $x_1,\cdots,x_n$ 的 Vandermonde 矩阵　2.6.1//162

$i(A)\equiv(i_+(A),i_-(A),i_0(A))$　$n\times n$ 矩阵 A 的惯性　3.1.2//171,

　　3.14.2//228

$\otimes$　Kronecker 乘号　7.1.1//427

vec$A$　矩阵 A 的相伴向量　7.1.4//428

A^I　广义逆　7.5.2//468

A^+　Moore-Penrose 逆　7.6.2//473

A^-　(1)-逆　7.7.1//484

A^D　Drazin 逆　7.8.1//489

deg $p(t)$，deg $A(\lambda)$　多项式 $p(t)$ 的次数、λ-矩阵 $A(\lambda)$ 的次数

　　9.6.1//601

参考文献

[A1975] S. S. Agaian, *Hadamard matrices and their applications*, Springer-Verlag, Berlin, 1975.

[A1956] A. C. Aitken, *Determinants and matrices*, 9th ed., Oliver and Boyd, Edinburgh, 1956.

[AH1983] G. Alefeld and J. Herzberger, *Introduction to interval computations*, Academic Press, New York, 1983.

[A1998] A. L. Andrew, *Centrosymmetric matrices*, SIAM Rev., 40(1998), 697-698.

[AGZ1956] R. J. Arms, L. D. Gates and B. Zondek, *A method of block iteration*, J. Soc. Indust. Appl. Math., 4(1956), 220-229.

[B1971] S. Barnett, *Matrices in control theory with applications to linear programming*, Van Nostrand Reinhold, London, 1971.

[B1979] S. Barnett, *Matrix methods for engineers and scientists*, McGraw-Hill, London, 1979.

[B1990] S. Barnett, *Matrix: methods and applications*, Clarendon Press, Oxford, 1990.

[B1984] S. Barnett, *Matrices in control theory*, 2nd ed., Krieger, Florida, 1984.

[B1983] A. Basilevsky, *Applied matrix algebra in the statistical sciences*, North-Holland, New York, 1983.

[B1965] E. F. Beckenbach and R. Bellman, *Inequalities*, Springer-Verlag, New York, 1965.

[B1988] 北京大学数学系几何与代数教研室代数小组编，高等代数，高等教育出版社，北京，1988.

[BL1988] G. R. Belitskii and Y. I. Lyubich, *Matrix norms and their applications*, Birkhäuser, Basel, 1988.

[B1940] E. T. Bell, *The development of mathematics*, McGraw Hill, New York, 1940.

[B1960] R. Bellman, *Introduction to matrix analysis*, McGraw-Hill, New York, 1960.

[BG1980] A. Ben-Israel and T. N. E. Greville, *Generalized inverses: theory and applications*, 2nd edn., Krieger, Florida, 1980.

[BP1979] A. Berman and R. Plemmons, *Nonnegative matrices in the mathematical sciences*, Academic Press, New York, 1979.

[B1997] R. Bhatia, *Matrix analysis*, Springer, New York, 1997.

[C1929] F. Caajori, *A history of mathematical notations*, Vol. 2, Open Court Publishing Company, Chicago, 1929.

[CH1966] P. Camion and A. J. Hoffman, *On the nonsingular of complex matri-ces*, Pacific J. Math., 17(1966), 211-214.

[CCG1997] R. H. Chan, T. F. Chan and G. H. Golub (eds), *Iterative methods in scientific computing*, Springer-Verlag, Singapore, 1997.

[C1998] H. C. Chen, *Generalized reflexive matrices: special properties and applications*, SIAM J. Matrix Anal. Appl., 19(1998), 140-153.

[C1987] 陈景良, 近代分析数学概要, 清华大学出版社, 北京, 1987.

[D1979] P. J. Davis, *Circulant matrices*, Wiley, New York, 1979.

[E1995] H. M. Edwards, *Linear algebra*, Birkhäuser, Boston, 1995.

[F1986] M. Fiedler, *Special matrices and their applications in numerical mathematics*, Martinus Nijhoff, Dordrecht, 1986.

[GG1995] J. Gilbert, and L. Gilbert, *Linear algebra and matrix theory*, Acade-mic Press, San Diego, 1995.

[GLR1982] I. Gohberg, P. Lancaster, and L. Rodman, *Matrix polynomials*, Aca-demic Press, New York, 1982.

[GV1989] G. Golub and C.VanLoan, *Matrix computations*, 2nd edn., Johns Hopkins University Press, Baltimore, 1989.

[G1981] A. Graham, *Kronecker products and matrix calculus with applications*, Horwood, Chichester, U.K., 1981.

[G1983] F. A. Graybill, *Matrices with applications to statistics*, 2d ed., Wad-sworth, Belmont, Calif., 1983.

[GC1990] 关治, 陈景良, 数值计算方法, 清华大学出版社, 北京, 1990.

[G1988] 郭本瑜, 偏微分方程的差分方法, 科学出版社, 北京, 1988.

[H1996] A. S. Hadi, *Matrix algebra as a tool*, Duxbury Press, Belmont, Calif, 1996.

[H1997] D. A. Harville, *Matrix algebra from a statistician's perspective*, Springer, New York, 1997.

[HS1979] M. Harwit and N.J. A. Sloane, *Hadamard transform optics*, Academic Press, New York, 1979.

[HR1984] G.Heinig and R. Rost, Algebraic methods for Toeplitz-like matrices and operators, Akademie-Verlag, Berlin, 1984.

[HW1988] I. N. Herstein and D. J. Winter, *Matrix theory and linear algebra*, Macmillan, New York, 1988.

[H1986] R. Hill, *A first course in coding theory*, Clarendon Press, Oxford, 1986.

[HJ1985] R. A. Horn and C. R. Johnson, *Matrix analysis*, Cambridge University Press, Cambridge, 1985.

[HJ1991] R. A. Horn and C. R. Johnson, *Topics in matrix analysis*, Cambridge University Press, Cambridge, 1991.

[H1963] P. Horst, *Matrix algebra for social scientists*, Holt, Rinehart and Winston, New York, 1963.

[H1964] A. S. Householder, *The theory of matrices in numerical analysis*, Blaisdell, New York, 1964.

[H1991] 胡家赣, 线性代数方程组的迭代解法, 科学出版社, 北京, 1991.

[JM1992] A. Jennings, and J. J. McKeown, *Matrix computation*, John Wiley, Chichester, West Sussex, New York, 1992.

[J1990] C. R. Johnson, ed., *Matrix theory and applications*, Proceeding of Applied Mathematics, vol. 40, American Mathematical Society, Providence, R. I., 1990.

[K1982] K. H. Kim, *Boolean matrix theory and applications*, Marcel Dekker, New York, 1982.

[LT1985] P. Lancaster and M. Tismenetsky, *The theory of matrix with applications*, 2nd edn., Academic Press, New York, 1985.

[L1997] P. D. Lax , *Linear algebra*, Wiley, New York, 1997.

[L1994] D. C. Lay, *Linear Algebra and its applications*, Addison-Wesley, Reading, Massachusetts, 1994.

[MS1977] F. J. MacWilliams and N. J. A. Sloane, *The theory of error-correcting codes*, North-Holland, Amsterdam. 1977.

[MM1964] M. Marcus and H. Minc, *A survey of matrix theory and matrix inequalities*, Allyn and Bacon, Boston, 1964.

[M1936] N. H. McCoy, *On the characteristic roots of matrix polynomials*, Bull. Amer. Math. Soc. 42(1936), 592-600.

[M1988] H. Minc, *Nonnegative matrices*, Wiley, New York, 1988.

[MT1952] T. S. Motzkin and O. Taussky, *Pairs of matrices with property L*, Trans. Amer. Math. Soc., 73(1952).

[N1970] E. D. Nering, *Linear algebra and matrix theory*, Wiley, New York. 1970.

[N1972] M. Newman, *Integral matrices*, Academic Press, New York, 1972.

[O1960] A. M. Ostrowski, *Solution of equations and systems of equations*, Academic Press, New York and London, 1960.

[P1976] N. J. Pullman, *Matrix theory and its applications: selected topics*, M. Dekker, New York, 1976.

[Q1990] 清华大学应用数学系《现代应用数学手册》编委会,现代应用数学手册: 计算方法分册, 北京出版社, 北京, 1990.

[Q1998] 清华大学应用数学系《现代应用数学手册》编委会,现代应用数学手册: 应用分析分册, 清华大学出版社, 北京, 1998.

[RM1971] C. R. Rao and S. K. Mitra, *Generalized inverse of matrices and its applications*, John Wiley, New York, 1971.

[RR1998] C. R. Rao and M. B. Rao, *Matrix algebra and its application to statistics and econometrics*, World Scientific, Singapore, 1998.

[R1968] A. Rogers, *Matrix analysis of interregional population growth and distribution*, University of California Press, Berkeley, 1968.

[S1997] J. R. Schott, *Matrix analysis for statistics*, John Wiley, New York, 1997.

[S1986] A. Schrijver, *Theory of linear and integer programming*, Wiley, Chichester, 1986.

[S1966] S. R. Searle, *Matrix algebra for the biological sciences, including application in statistics*, Wiley, New York, 1966.

[S1970] S. R. Searle, *Matrix algebra for business and economics*, Wiley-Interscience, and W. H. Hausman, New York, 1970.

[S1981] E. Seneta, *Nonnegative matrices*, Springer-Verlag, New York, 1981.

[ST1968] D. A. Suprenenko and R. I. Tyshkevich, *Commutative matrices*, Academic Press, New York, 1968.

[T1964] J. F. Traub, *Iterative.methods for the solution of equations*, Prentice-Hall, New Jersey, 1964.

[T1950] H. W. Turnbull, *The theory of determinants, matrices, and invariants*, Blackie, London and Glasgow, 1950.

[TA1932] H. W.'Turnbull, and A. C. Aitken, *An introduction to the theory of canonical matrices*, Blackie, London and Glasgow, 1932.

[V1993] R. J. Valenza, *Linear algebra, an introduction to abstract mathema-tics*, Springer-Verlag, 1993.

[V1962] R. S. Varga, *Matrix iterative analysis*, Prentice-Hall, Englewood Cliffs, N. J., 1962.

[W1985] J. R. Weaver, *Centrosymmetric (cross-symmetric) matrices, their basic properties, eigenvalues, and eigenvectors*, Amer. Math. Monthly, 92(1985), 711-717.

[WG1980] W. Weaver, and J. M. Gere, Matrix analysis of framed structures, D. Van Nostrand, New York, Toronto, 1980.

[W1968] J. R. Westlake, *A handbook of numerical matrix inversion and solution of linear equations*, Wiley, New York, 1968.

[W1965] J. H. Wilkinson, *The algebraic eigenvalue problem*, Clarendon Press, Oxford, 1965.

[X1995] 徐树方，矩阵计算的理论和方法，北京大学出版社，北京，1995.

[Y1954] D. M. Young, *Iterative method for solving partial difference equations of elliptic type*, Trans. Amer. Math. Soc., 76(1954), 92-111.

[Z1999] F. Zhang, *Matrix theory: basic resultes and techniqus*, Springer, New York, Berlin, 1999.

索　引

术 语　　子 节 号//页 码

D

E

F

G

H

I

J

K

L

M

N

O

P

T

U

V

W

X

Y